UNIVERSITÀ POLITECNICA DELLE MARCHE
Repository ISTITUZIONALE

Investigating the phenomenon of NSFW posts in Reddit

(Article begins on next page)

13 March 2025

Investigating the phenomenon of NSFW posts in Reddit

Enrico Corradini[1], Antonino Nocera[2*], Domenico Ursino[1], Luca Virgili[1]

[1] Department of Information Engineering, Polytechnic University of Marche,
[2] Department of Electrical, Computer and Biomedical Engineering, University of Pavia,
* Corresponding author
{e.corradini, l.virgili}@pm.univpm.it; antonino.nocera@unipv.it; d.ursino@univpm.it

Abstract

In this paper, we study the characteristics of NSFW (Not Safe For Work) posts in Reddit, highlighting their differences from SFW (Safe For Work) posts, which have been much more studied in the past literature. In our investigation, we studied all Reddit posts from 2019. Through both descriptive analytics techniques and social network analysis techniques, we extract three findings on the main differences between NSFW and SFW posts in Reddit. Thanks to these findings, we are able to better understand the dynamics (authors, subreddits, readers) behind NSFW posts. In particular, it becomes clear that this is a niche world where authors are strongly cohesive. However, at the same time, the most popular ones show a clear opening to new authors, whom they are willing to collaborate with, from the beginning.

Keywords: Reddit; NSFW Posts; NSFW Authors; Co-posting Authors; Assortativity of NSFW Authors; Knowledge Extraction

1 Introduction

Reddit[1] is currently one of the most active social media. It has been extensively studied by researchers in the past [19]. In [25], the authors present an interesting longitudinal analysis of the evolution of this social medium. Furthermore, many papers have focused on specific aspects of this social network, concerning, for example, community structures and interactions [27, 8, 10], user behavior [3, 14, 16], structure and content of subreddits, posts and comments [24], structural properties [10, 13, 33], text classification [15], user migration [21], political and ideological aspects [12, 31].

One aspect of Reddit worth to be analyzed involves NSFW (Not Safe For Work) posts. This term refers to user-submitted content not suitable to be viewed in public or in professional contexts. The phenomenon of NSFW posts in Reddit has been very little investigated, although it is very common in this social medium. In fact, only a very small number of authors have analyzed it [17, 20]. The term "NSFW" has been proposed since 1998, and is one of the oldest acronyms of the Internet. Since its first appearance, many social media, such as Twitter, WhatsApp and Reddit, have adopted it to

---

[1] https://www.reddit.com

indicate certain sections or contents. In addition, several authors have focused on the analysis of this phenomenon in other social networks. The study about the role of images and selfies in NSFW content of `tumblr.com`, presented in [28], and the analysis of the anonymity level of NSFW content in both Twitter and Whisper, described in [7] are two examples.

In this paper, we give a contribution in this setting investigating the phenomenon of NSFW posts in Reddit and describing the whole context (authors, subreddits and readers) behind it. For this purpose, we consider a dataset that includes all the posts published in Reddit from January 1st, 2019 to December 31st, 2019.

During our investigation, we carried out three types of analysis, namely:

- *Descriptive Analysis*, to study the distributions of the entities involved in the phenomenon (e.g., the distribution of NSFW posts against subreddits, authors, score and comments).

- *Social Network Analysis*, to study the co-posting phenomenon, and therefore the interactions between authors of NSFW posts.

- *Assortativity Analysis*, to extend and deepen the previous analyses to discover and study whether possible forms of assortativity [22] exist among the authors of NSFW posts. Recall that assortativity is a particular case of homophily in social networks [18], which indicates the tendency of a node to cooperate with nodes having similar characteristics.

These analyses allowed us to extract three findings regarding NSFW posts, NSFW authors and NSFW subreddits, respectively. Throughout our analysis, in most of the cases, we compare each finding on NSFW posts with the corresponding one on SFW (Safe For Work) posts. Some of the questions these findings provide an answer to are the following:

- What can be said about the spread of NSFW posts in the subreddits?

- What can be said about the quantity of posts an NSFW author usually submits?

- What can be said about the score of NSFW posts?

- What can be said about the number and the score of comments to NSFW posts?

- What can be said about the level of interconnection between authors of NSFW posts?

- Is there a backbone among experienced authors of NSFW posts? In other words, do they tend to interact only with their peers (i.e., authors with the same level of experience), or are they open to collaborations with new authors who have just started publishing NSFW posts?

Finally, we suitably combine the knowledge represented by the three findings in order to describe the dynamics behind the phenomenon of NSFW posts in Reddit.

The rest of this paper is organized as follows: In Section 2, we present related literature. In Section 3, we describe the dataset used in our analysis. In Section 4, we provide an overview of our investigation activity. In Section 5, we study various distributions involving NSFW posts. In Section 6, we study several distributions regarding comments of NSFW posts. In Section 7, we investigate the

2

co-posting activity of the authors of NSFW posts. In Section 8, we evaluate the assortativity of the authors of NSFW posts. In Section 9, we combine the three findings derived during our investigations in order to define an overall picture of this phenomenon. Finally, in Section 10, we draw our conclusions and think of some possible developments of our research efforts.

2 Related literature

The term "NSFW" was first proposed in 1998 and it is one of the oldest acronyms of the Internet. It refers to content that is not suitable to be viewed in a working environment. Since then, different online systems, like Twitter, WhatsApp, many forums, and Reddit, have adopted this term to label sections with posted content not adequate for everybody and, in general, not suitable for public and professional contexts. Specifically, Reddit has introduced a dedicated group of contents called NSFW to separate posts suitable to be enjoyed in any context from those that should be watched in private environments.

Even if the contents of NSFW posts are considered side-contents to be kept separated from front-end contents, several researchers have started to study the characteristics of these contents, as well as the communities underneath them [17, 4, 28, 32, 7, 11].

From a high level analysis of the research efforts in the context of NSFW content, we may distinguish two main directions. The former focuses on understanding the main characteristics of people publishing or viewing such materials, as well as the features of the NSFW content itself. The latter, instead, uses features of NSFW content to build content detection and filtering solutions, often with the objective of enabling/disabling the visualization of this material for users.

In particular, the work described in [28] is an example of the first research direction. Here, the author investigates the role of images and selfies in NSFW contents of `tumblr.com`. NSFW contents, having the explicit NSFW label assigned by their authors, are extracted from tumblr blogs. Then, the described analysis focuses on images and reactions (interactions) surrounding them. The aim of this study is understanding the different roles that people assign to images and selfies, leading to the creation (or breaking) of trust relationships between users. Furthermore, the author provides evidence that different opinions about the membership of an image to the NSFW category may lead to violations of assumed trust between two individuals, thus causing the dissolution of a community.

Another contribution in the first research direction is the one reported in [7]. In this paper, the authors try to understand both the nature of the content posted in anonymous social media and the difference between NSFW content posted in these media and in non-anonymous ones (like, e.g., Twitter). To do this, they define an anonymity sensitivity metrics measuring how much users think that a post should be anonymous. Then, they use this metrics, in conjunction with a human annotator, to identify NSFW posts with the same level of anonymity sensitivity in Whisper (an anonymous media system) and Twitter. Hence, they carry out a deep comparative analysis of the two sets of posts and find that, actually, there is a strong difference between them, especially when it comes to the shades or levels of anonymity and their linguistic features.

Even if its main focus is slightly different from the one defining this first research direction, the work described in [17] gives a mentionable contribution in this setting. Indeed, the author considers a particular protest carried out by moderators of Reddit in 2015, when participants disabled their

subreddits to block posting activities. In this context, the author studies the different behavior of NSFW and SFW subreddit moderators. The results show that, even if several confounding factors could be considered to understand the underlying dynamics, NSFW subreddit moderators were more inclined to join the protest and block posting activities.

In the second research direction mentioned above, several works have been published in recent scientific literature [20, 9, 2, 32, 6]. For instance, the work described in [20] focuses on the protection of minors accessing the Internet from the exposure to unwanted and harmful contents. The proposed system can be seen as both an active content filtering solution, which protects the access of minor users to NSFW content, and a watchdog constantly monitoring and moderating websites to avoid the diffusion of unwanted content.

The problem of classifying video content as NSFW is faced in [9]. In this paper, the authors exploit Convolutional Neural Networks (ConvNets) for extracting audio-video patterns from NSFW videos. Specifically, they first extract separated audio and video features and then merge the two feature sets to obtain a single feature vector. After that, they provide this vector in input to some baseline classifiers. Even if the approach is naive, the achieved results outperform those of other methods, thus proving the adequacy of this proposal.

Similarly, the approach of [2] makes use of a deep neural network based solution to identify content belonging to the NSFW category. This approach is based on a residual network, which returns a value specifying the probability that a given content belongs to NSFW category. Moreover, it allows the computation of the degree of explicitness of the analyzed content, which can be used to feed a filtering system. Finally, it is capable of labeling media content with tampered extension to warn users about the potential risk of suspiciously unwanted material. The experiments show very interesting performance for this approach, which reaches an accuracy of about 96% also on image and video contents.

Still in this context, also the approach described in [32] makes use of a fast Convolutional Neural Network (CNN) for the detection of both NSFW and SFW images. Specifically, this proposal deals with the design of a neural network based solution to detect pictures with nudity in NSFW contents. After that, it defines picture filtering strategies for online media services.

Finally, the approach described in [6] strives to build a classifier for detecting NSFW content by looking at images and visual material in the post. The proposed solution uses a weighted sum of the results of multiple deep neural network models. The weighted combination is obtained by learning a linear regression model through Ordinary Least Squares. The authors prove that their solution outperforms the state of the art solutions based on single CNN models. For this purpose, they present a deep comparison on a manual labeled dataset.

Our approach is somehow near to the studies belonging to the first research direction introduced above. However, these approaches only study the content of NSFW posts and none of them focus on the structural network-based properties of NSFW and SFW posts and authors. Instead, we want to study such differences between the two categories with a comparative approach and typical Social Network Analysis methodologies.

The identified findings can be fundamental to improve existing techniques for content detection, parental control or content filtering solutions, such as the ones mentioned above. To the best of our knowledge, no similar studies have been conducted in social media platforms. Our paper aims at

providing a first contribution in this setting using Reddit as reference social network. However, as we will see below, our investigation strategy is general and can be specialized to other social media [5].

3 Dataset description

The dataset used for our analysis has been downloaded from the website `pushshift.io` [1], one of the main Reddit data sources. In particular, we extracted all the posts published on Reddit from January 1st, 2019 to September 1st, 2019[2]. The number of posts available for our analysis was 150,795,895. In Reddit, an NSFW post must be marked as such by its author. Therefore, there is no need for automatic labeling by Reddit or manual labeling by third-parties. If the user specifies that a post she/he is publishing is NSFW, Reddit puts a red label when displaying it and sets the value of the `over_18` field in its database to `true`. We used the value of this field to separate NSFW posts from SFW ones in our analyses.

We performed a preliminary ETL (Extraction, Transformation and Loading) activity on our dataset. In Data Analytics, this activity is typically carried out prior to any data analysis campaign. It aims at cleaning the data in the dataset, removing any errors and inconsistencies, integrating any data from different sources, and transforming the cleaned and integrated data into a single format chosen for the next data analysis tasks [23].

During the ETL phase, we observed that some of the available posts were made by authors who had left Reddit. We decided to remove these posts from our dataset. At the end of this activity, the number of available posts was 122,568,630. NSFW posts were 11,908,377, equivalent to 9.72% of them.

As pointed out in the Introduction, the goal of our paper is to understand the characteristics of NSFW posts and their authors, comparing them with the SFW posts and their authors. For this reason, we decided to extract from the dataset described above two sub-datasets, with the same number of posts each. Both of them are limited to January and February 2019. The first dataset $\mathcal{D}$ contains only SFW posts, while the second, called $\overline{\mathcal{D}}$, stores only NSFW posts. We randomly selected 1,250,000 posts for each of them to reduce the datasets' size and the computation time. It should be noted that this number is absolutely in line with the number of posts generally used in the analyses of Reddit [29, 21, 26, 12]. However, we repeated all the analyses on two other datasets $\mathcal{D}'$ and $\overline{\mathcal{D}}'$ to verify the stability of our results. The set $\mathcal{D}'$ (resp., $\overline{\mathcal{D}}'$) consists of 1,250,000 SFW (resp., NSFW) posts published in March and April 2019, randomly selected from the original dataset. In addition, we carried out a deeper stability check evaluating all posts of 2019 month by month (see Section 6.4).

As a preliminary analysis, we focused on the "context" of SFW and NSFW posts. Here, we use the term "context" of a post to denote its author, its comments and the subreddits in which it was published. In this analysis, we wanted to verify if the context of SFW posts and the one of NSFW posts are the same or not. To answer this question, we calculated the values of some parameters on $\mathcal{D}$ and $\overline{\mathcal{D}}$ and, then, on $\mathcal{D}'$ and $\overline{\mathcal{D}}'$. The results obtained are shown in Table 1.

This table shows that the reference contexts for SFW and NSFW posts are basically independent.

---

[2]Actually, only for stability analysis, we considered all the posts from January 1st, 2019 to December 31st, 2019 (see Section 6.4).

Parameter	$\mathcal{D}$ and $\overline{\mathcal{D}}$	$\mathcal{D}'$ and $\overline{\mathcal{D}}'$
Number of authors who published at least one SFW post	59,465	58,561
Number of authors who published only SFW posts	58,801	57,891
Percentage of authors publishing SFW posts who published only posts of this type	98.88%	98.52%
Number of authors who published at least one NSFW post	36,758	36,461
Number of authors who published only NSFW posts	36,094	36,131
Percentage of authors publishing NSFW posts who published only posts of this type	98,19%	99.09%
Number of subreddits containing at least one SFW post	89,360	92,445
Number of subreddits containing only SFW posts	82,050	85,157
Percentage of subreddits containing SFW posts that contain only posts of this type	91.82%	92.12%
Number of subreddits containing at least one NSFW post	41,365	45,910
Number of subreddits containing only NSFW posts	34,055	38,622
Percentage of subreddits containing NSFW posts that contain only posts of this type	82.33%	84.13%

Table 1: Parameters about the authors and the subreddits of SFW and NSFW posts - $\mathcal{D}$ (resp., $\overline{\mathcal{D}}$) stores SFW (resp., NSFW) posts of January and February 2019, while $\mathcal{D}'$ (resp., $\overline{\mathcal{D}}'$) stores the same kind of post but for March and April 2019

In fact, more than 98% of authors writing SFW posts do not write NSFW posts, and vice versa. In addition, more than 91% of subreddits containing SFW posts do not contain NSFW posts, and more than 82% of subreddits containing NSFW posts do not contain SFW posts. Another important result is that all the computations are stable over time because the values obtained for January and February 2019 (Jan-Feb, for short) are very similar to the ones returned for March and April 2019 (Mar-Apr, for short).

4   Overview of our investigation activity

Our investigation of the phenomenon of NSFW posts in Reddit follows the workflow shown in Figure 1.
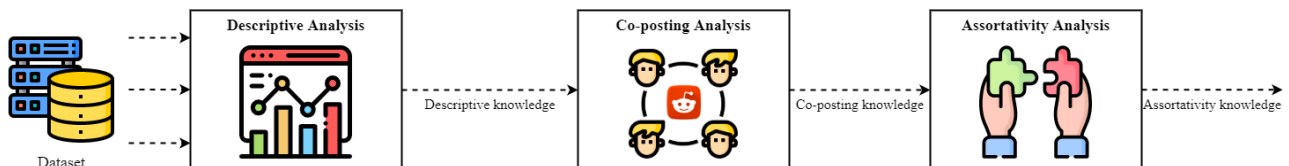

Figure 1: The workflow representing the tasks of our investigation

Due to layout reasons, this figure shows the dataset in input only to the first module. Actually, the dataset is provided in input to each module of the workflow. Similarly, the descriptive (resp., co-posting) knowledge, which are shown as an input for the co-posting (resp., assortativity) analysis module, are also an output of the investigation activity.

As we can see in Figure 1, the first phase of our investigation consists of a descriptive analysis of the phenomenon of NSFW posts in Reddit. In particular, this analysis extracts knowledge through several distributions involved in the phenomenon (e.g., the distribution of NSFW posts against subreddits, authors and scores, the distribution of the authors of NSFW posts against subreddits, etc.).

This knowledge, together with the original dataset, represents the input of the second phase of our investigation. This analysis employs a social network derived from co-posting activities of authors. In particular, the nodes of this social network represent the authors, whereas the edges indicate co-posting activities between them. Starting from this social network, we study the co-posting activities

of the authors of SFW and NSFW posts and extract the corresponding knowledge.

The result of this analysis, together with the original dataset, represents the input of the third phase. It leverages the social network built in the previous phase to carry out analyses on the assortativity of the authors of SFW and NSFW posts with respect to some forms of centrality. The results obtained during this phase represent the last kind of knowledge returned as output by our approach.

5 Investigating distributions involving NSFW posts

In this section, we present some analyses directly involving NSFW and SFW posts. In particular, we study the distribution of subreddits and authors against posts and the distribution of posts against the scores assigned to them by Reddit users.

5.1 Distribution of subreddits against posts

We computed the distributions of the subreddits against NSFW and SFW posts for the datasets $\mathcal{D}$ and $\overline{\mathcal{D}}$. The results obtained are reported in Figure 2.

This figure shows that the two distributions follow a power law. We also computed some parameters for the two power law distributions; they are shown in the second and third columns of Table 2. To verify the stability of results found, we made the same computations on $\mathcal{D}'$ and $\overline{\mathcal{D}}'$ datasets. They are shown in the fourth and fifth columns of Table 2.

Parameter	SFW posts Jan-Feb	NSFW posts Jan-Feb	SFW posts Mar-Apr	NSFW posts Mar-Apr
Maximum number of subreddits	47,480 (53.13%)	18,332 (44.31%)	49,502 (53.24%)	21,034 (45.02%)
Number of subreddits of the 99 percentile	1,095	571	1,101	569
Maximum number of posts	25,006 (4.62%)	34,424 (4.57%)	26,650 (4.98%)	31,329 (4.76%)
Number of posts of the 99 percentile	7,719	9,862	7,721	9,859
Average number of subreddits	126	54	137	57
Average number of posts	767	981	768	905
α (power law parameter)	1.6539	1.6974	1.6767	1.6859
δ (power law parameter)	0.0266	0.0364	0.0306	0.0432

Table 2: Parameters of the distributions of subreddits against posts

From this table, we can observe that the maximum and the average numbers of subreddits for SFW posts is more than twice the value obtained for NSFW posts. The maximum and the average numbers of NSFW posts in a subreddit are slightly higher than SFW posts. There are no significant differences in the α and δ parameters of the two power law distributions. Indeed, both of them are very steep. The comparison of the second and the third columns of Tables 2, on the one hand, and the fourth and fifth columns of the same table, on the other hand, also tells us that the trends obtained are stable over time, because their variations between Jan-Feb and Mar-Apr are not significant.

Although the two curves show almost identical trends, as confirmed by the similar values of α and δ, we found interesting the differences in the maximum and average values. In other words, the curve shapes are similar but the ranges of values are different. To confirm these results we compared the two distributions through the *Wilcoxon rank sum test* [30].

This test indicated that the number of subreddits in which Jan-Feb SFW posts were published was statistically significantly higher than the corresponding one of NSFW posts ($\tau = 2.8 \cdot 10^{-4}, p < 0.01$).
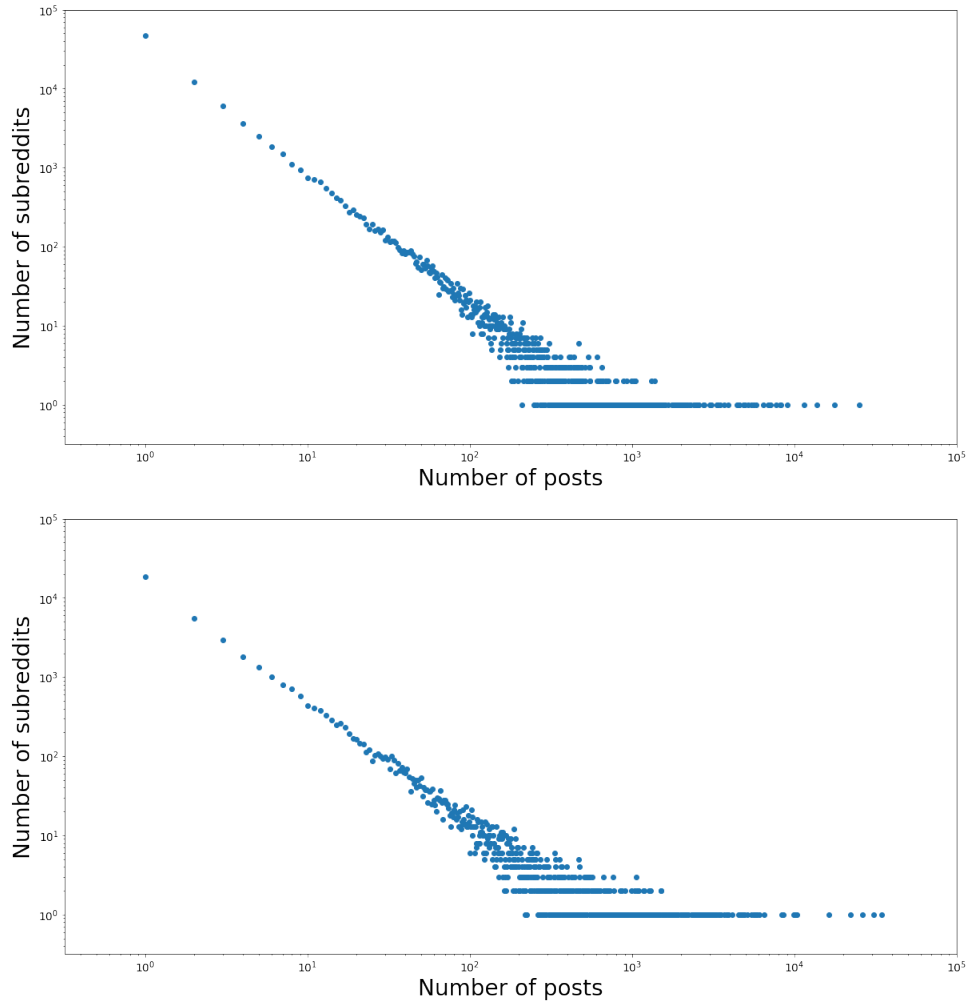
Figure 2: Log-log plots of the distributions of subreddits against SFW posts (on top) and NSFW posts (on bottom) - Datasets regarding January and February 2019

This result can be explained taking into account the intrinsic nature of NSFW posts, whose content is certainly less suitable for the general public than the one of SFW posts.

5.2 Distribution of authors against posts

Figure 3 shows the distributions of authors against SFW and NSFW posts for the datasets $\mathcal{D}$ and $\overline{\mathcal{D}}$. From the analysis of this figure we can see that both distributions follow a power law.

In Table 3, we report the main parameters of these two power law distributions for the datasets $\mathcal{D}$ and $\overline{\mathcal{D}}$, on one hand, and $\mathcal{D}'$ and $\overline{\mathcal{D}}'$, on the other hand.

A Wilcoxon rank sum test showed that the number of authors of Jan-Feb SFW posts was statistically significantly higher than the corresponding one of NSFW posts ($\tau = 1.2 \cdot 10^{-4}, p < 0.01$).

This result can also be explained taking into account the topics of NSFW posts. Indeed, these are more specific than those involving SFW posts. Differently from SFW posts that can be written by
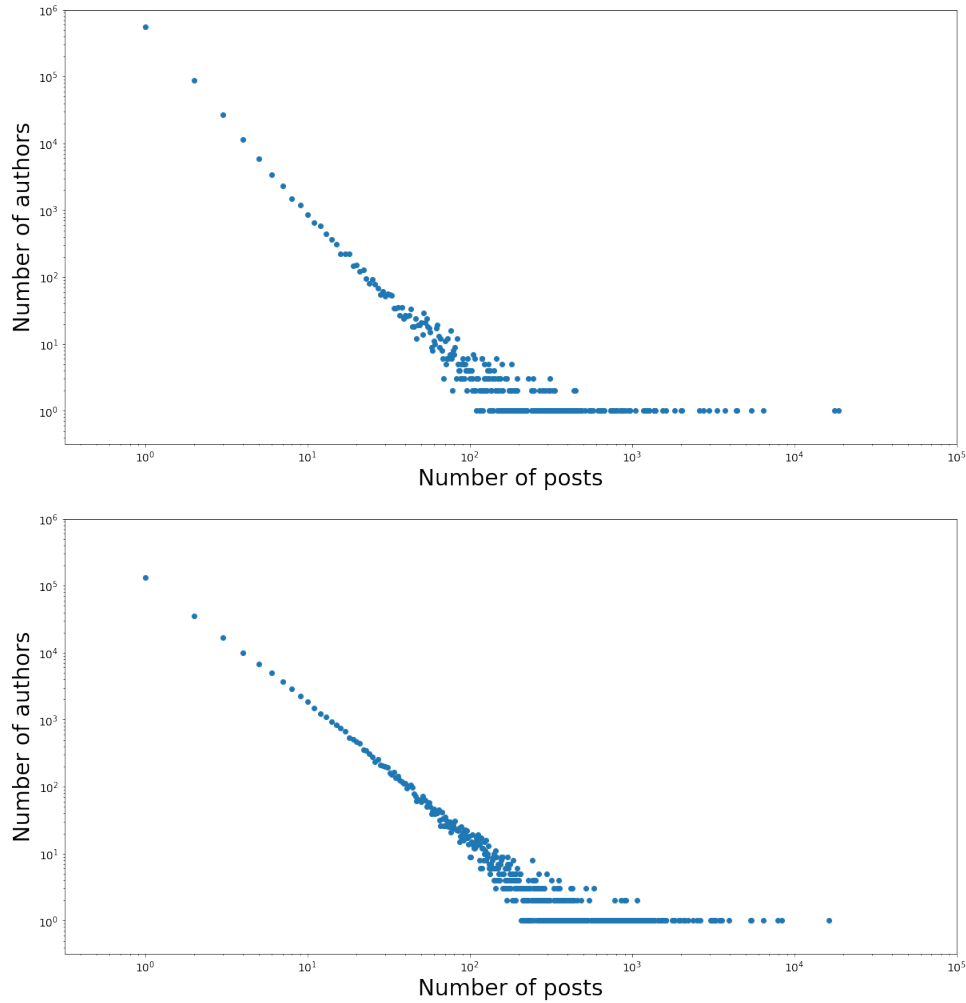
Figure 3: Log-log plots of the distributions of authors against SFW posts (on top) and NSFW posts (on bottom) - Datasets regarding January and February 2019

Parameter	SFW posts Jan-Feb	NSFW posts Jan-Feb	SFW posts Mar-Apr	NSFW posts Mar-Apr
Maximum number of authors	555,854 (79.06%)	131,070 (56.43%)	551,863 (78.97%)	133,594 (57.01%)
Number of authors of the 99 percentile	11,471	5,055	11,469	5,052
Maximum number of posts	18,724 (11.85%)	16,383 (5.70%)	16,513 (10.98%)	15,674 (5.48%)
Number of posts of the 99 percentile	5,426	5,393	5,424	5,393
Average number of authors	2,190	439	2,083	416
Average number of posts	491	543	491	521
α (power law parameter)	1.4631	1.5566	1.4505	1.5435
δ (power law parameter)	0.0473	0.0353	0.0304	0.0287

Table 3: Parameters of the distributions of authors against posts

anyone, the authors who generally publish NSFW posts are a small circle of people almost exclusively dedicated to this type of post. Consequently, while it is true that NSFW posts are much fewer than SFW posts, it is also true that they are published by an extremely limited number of authors. This explains the result.

5.3 Distribution of posts against scores

A newly submitted post on Reddit has a score of 1. A user can upvote (resp., downvote) the post, increasing (resp., decreasing) its score by 1. We have computed the distributions of SFW and NSFW posts against scores for the datasets $\mathcal{D}$ and $\overline{\mathcal{D}}$, and, then, for $\mathcal{D}'$ and $\overline{\mathcal{D}}'$, on the other hand. For the sake of simplicity, in Table 4, we report the main parameters of these distributions, which again follow a power law.

Parameter	SFW posts Jan-Feb	NSFW posts Jan-Feb	SFW posts Mar-Apr	NSFW posts Mar-Apr
Maximum score	183,453 (57.98%)	106,947 (47.26%)	191,864 (61.87%)	112,830 (49.62%)
Number of score of the 99 percentile	4,746	3,645	4,825	3,275
Average score	9,881	4,191	8,809	3,819
α (power law parameter)	1.5998	1.5140	1.6061	1.5165
δ (power law parameter)	0.0197	0.0366	0.0154	0.0355

Table 4: Parameters of the distributions of posts against scores

A Wilcoxon rank sum test showed that the score of Jan-Feb SFW posts was statistically significantly higher than the corresponding one of NSFW posts ($\tau = 0.00109, p < 0.01$).

Once again, this result can be explained by the type of contents that generally characterizes NSFW posts.

5.4 Analysis of positive and negative posts for SFW and NSFW cases

In the previous section, we have observed that each post has a score, initially equal to 1, which can increase or decrease based on the upvotes or downvotes of users. Actually, Reddit does not report the posts with a negative score in its database. For this reason, the values of the scores both in Reddit and in `pushshift.io` range in the interval $[0, +\infty)$. In this setting, posts with a score equal to 0 are particularly relevant, because they are the only ones that have been rated negatively by at least one user, or have received more downvotes than upvotes.

We computed the distributions of authors against negative posts for both SFW and NSFW posts. In both cases, we have found that they follow a power law. We report the main parameters of these distributions in Table 5.

Parameter	SFW posts Jan-Feb	NSFW posts Jan-Feb	SFW posts Mar-Apr	NSFW posts Mar-Apr
Maximum number of authors	66,162 (92.31%)	24,607 (74.86%)	61,254 (91.98%)	24,172 (73.87%)
Number of authors of the 99 percentile	40,028	11,606	40,024	11,598
Maximum number of posts	133 (9.64%)	460 (14.38%)	103 (8.98%)	399 (13.76%)
Number of posts of the 99 percentile	126	369	122	370
Average number of authors	1,666	505	1,691	544
Average number of posts	32	49	28	47
α (power law parameter)	1.4360	1.4349	1.5512	1.4360
δ (power law parameter)	0.0615	0.0,0616	0.0543	0.0616

Table 5: Parameters of the distributions of authors against negative posts

A Wilcoxon rank sum test showed that the number of authors of Jan-Feb SFW negative posts was statistically significantly higher than the corresponding one of NSFW posts ($\tau = 5.1 \cdot 10^{-4}, p < 0.01$).

These conclusions, although interesting, must be intertwined with those regarding positive posts, to better characterize the features of negative ones. For this reason, we computed the distributions of

authors against positive posts. Also in this case, the distributions follow a power law similar to the previous ones. We report the values of the main parameters of these distributions in Table 6.

Parameter	SFW posts Jan-Feb	NSFW posts Jan-Feb	SFW posts Mar-Apr	NSFW posts Mar-Apr
Maximum number of authors	522,540 (79.66%)	124,054 (56.56%)	519,774 (79.54%)	126,602 (56.89%)
Number of authors of the 99 percentile	9,083	4,346	9,080	4,352
Maximum number of posts	18,684 (11.88%)	16,383 (5.77%)	16,481 (10.67%)	15,564 (5.73%)
Number of posts of the 99 percentile	5,165	4,638	5,160	4,641
Average number of authors	2,018	418	1,944	394
Average number of posts	483	541	493	514
α (power law parameter)	1.4318	1.5145	1.4855	1.5498
δ (power law parameter)	0.0311	0.0263	0.0275	0.0291

Table 6: Parameters of the distributions of authors against positive posts

A Wilcoxon rank sum test indicated that the number of authors of Jan-Feb SFW positive posts was statistically significantly higher than the corresponding one of NSFW posts ($\tau = 1.1 \cdot 10^{-4}, p < 0.01$).

We now compare Tables 5 and 6 to extract the features characterizing negative posts versus positive ones. There are no significant differences between positive and negative posts in the maximum and average number of authors of NSFW and SFW posts. The same is true for the average number of posts and the trends of the power law distributions. However, there is a very interesting aspect that differentiates negative posts from positive ones. Indeed, the maximum number of negative posts is much higher for NSFW posts than for SFW ones. This trend is not found in positive posts.

The explanation behind this result is the same as the one seen in Section 5.3.

5.5 Distribution of subreddits against authors

We computed the distributions of subreddits against the authors of SFW and NSFW posts. In both cases, we saw that they follow a power law similar to those shown in the previous figures. We report the values of the most important parameters in Table 7.

Parameter	SFW posts Jan-Feb	NSFW posts Jan-Feb	SFW posts Mar-Apr	NSFW posts Mar-Apr
Maximum number of subreddits	62,839 (70.32%)	29,798 (72.03%)	65,861 (71.12%)	33,963 (72.01%)
Number of subreddits of the 99 percentile	932	538	930	533
Average number of subreddits	151	87	161	101
Maximum number of authors	20,285 (5.70%)	11,161 (4.70%)	21,801 (5.64%)	11,326 (4,59%)
Number of authors of the 99 percentile	6,435	4,627	6,431	4,635
Average number of authors	604	499	601	481
α (power law parameter)	1.7143	1.7992	1.6944	1.7343
δ (power law parameter)	0.0302	0.0.0382	0.0288	0.0362

Table 7: Parameters of the distributions of subreddits against authors

A Wilcoxon rank sum test showed that: *(i)* the number of subreddits of Jan-Feb SFW posts was statistically significantly higher than the corresponding one of NSFW posts; *(ii)* the number of authors of Jan-Feb SFW posts was statistically significantly higher than the corresponding one of NSFW posts ($\tau = 6.3 \cdot 10^{-4}, p < 0.01$).

The explanation behind this result is essentially related to the fact that NSFW posts have particular contents that are of interest to a minority of people. Therefore, they are published in a limited number of subreddits.

In the next analyses, to save space, we will avoid highlighting those cases where the values α and δ of power law distributions are similar, as well as those cases where the parameter values are stable when switching from Jan-Feb to Mar-Apr. Only if one or both of these conditions are not valid in some analysis, we will explicitly highlight this situation.

6 Investigating distributions on comments to NSFW posts

In this section, we analyze the comments to NSFW posts investigating their authors, the scores they get and the subreddits they are submitted to.

6.1 Distribution of comments against posts

The distributions of comments against SFW posts and NSFW posts follow a power law. Table 8 shows the values of the main parameters of these distributions.

Parameter	SFW posts Jan-Feb	NSFW posts Jan-Feb	SFW posts Mar-Apr	NSFW posts Mar-Apr
Maximum number of posts	499,068 (2.29%)	667,942 (5.79%)	522,477 (2.94%)	676,606 (5.81%)
Number of posts of the 99 percentile	8,257	10,707	8,362	10,719
Maximum number of comments	41,478 (39.93%)	28,227 (53.43%)	36,283 (40.01%)	23,485 (51.32%)
Number of comments of the 99 percentile	10,582	21,983	9,985	22,735
Average number of comments	1,237	771	1,402	656
α (power law parameter)	1.4836	1.3990	1.4779	1.4353
δ (power law parameter)	0.0178	0.0304	0.0160	0.0291

Table 8: Parameters of the distributions of comments against posts

A Wilcoxon rank sum test showed that the number of comments of Jan-Feb SFW posts was statistically significantly higher than the corresponding one of NSFW posts ($\tau = 8.68 \cdot 10^{-5}, p < 0.01$).

As a further investigation on this topic, we considered both the top 150 most commented SFW and NSFW posts. As a first analysis, we observed that SFW (resp., NSFW) posts have been submitted by 141 (resp., 130) authors in 55 (resp., 77) different subreddits. This result highlights that there is no author or subreddit able to monopolize post comments. Indeed, the phenomenon is highly distributed.

Then, we computed the distributions of the number of these comments against subreddits. They are reported in Figure 4. Plots (a) and (b) of this figure show that the two distributions follow a power law. We computed the parameter values of these power laws and we obtained $\alpha = 3.41$ and $\delta = 0.075$ for SFW post comments, and $\alpha = 3.53$ and $\delta = 0.07$ for NSFW post comments. A Wilcoxon rank sum test indicated that the number of comments associated with the subreddits containing Jan-Feb SFW posts was statistically significantly higher than the corresponding one of NSFW posts ($\tau = 0.16493, p < 0.01$).

Finally, we computed the distribution of the number of these comments against authors. Also in this case, we found that it follows a power law. The values of the corresponding parameters are $\alpha = 3.06$ and $\delta = 0.03$ for SFW post comments and $\alpha = 2.20$ and $\delta = 0.03$ for NSFW post comments. The conclusions about the trend and the values are analogous to the previous ones.

A Wilcoxon rank sum test indicated that the number of comments for Jan-Feb SFW posts was statistically significantly higher than the corresponding one of NSFW posts ($\tau = 0.34951, p < 0.01$).

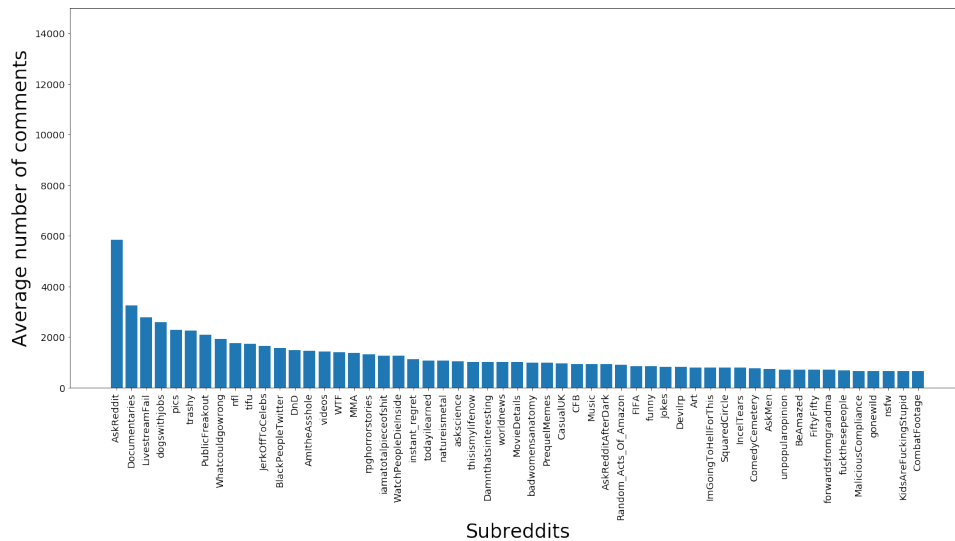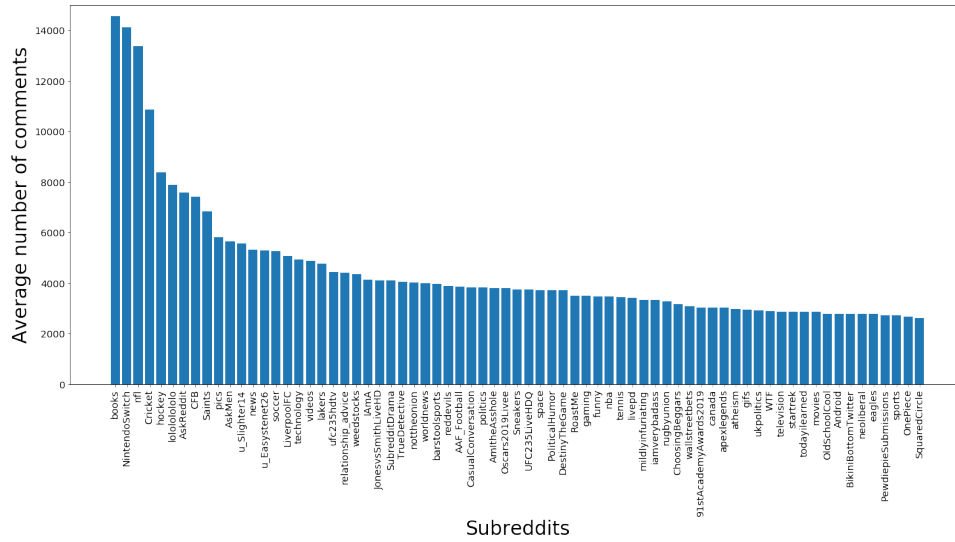The motivations behind this result are the same as those in Section 5.5.

Figure 4: Distributions of comments to the top 150 most commented SFW posts (on top) and NSFW posts (on bottom) against subreddits - Datasets regarding January and February 2019

6.2 Distribution of subreddits against comments

We computed the distributions of subreddits against the comments to SFW and NSFW posts. In both cases we obtained that they follow a power law and show trends similar to those shown in the previous figures. The main parameters of these distributions are reported in Table 9.

A Wilcoxon rank sum test showed that the number of comments associated with the subreddits containing Jan-Feb SFW posts was statistically significantly higher than the corresponding one of NSFW posts ($\tau = 6.34 \cdot 10^{-6}, p < 0.01$).

Once again, the motivations behind this result are the same as those in Section 5.5.

Parameter	SFW posts Jan-Feb	NSFW posts Jan-Feb	SFW posts Mar-Apr	NSFW posts Mar-Apr
Maximum number of comments	484,792 (5.45%)	301,040 (9.17%)	462,415 (5.41%)	244,912 (9.73%)
Number of comments of the 99 percentile	47,590	25,056	47,698	28,635
Average number of comments	3,942	2,607	3,800	2,391
α (power law parameter)	1.8025	1.7659	1.7981	1.7507
δ (power law parameter)	0.0236	0.0235	0.0217	0.0310

Table 9: Parameters of the distributions of subreddits against comments

6.3 Distribution of comments against scores

We computed the distributions of comments to SFW and NSFW posts against scores. They are reported in Figures 5 and 6 for the datasets $\mathcal{D}$ and $\overline{\mathcal{D}}$. These figures show that the corresponding distributions do not follow a power law, and this is the first case. As we can see from figures, the distributions are irregular, even if both of them seem having a Gaussian trend.
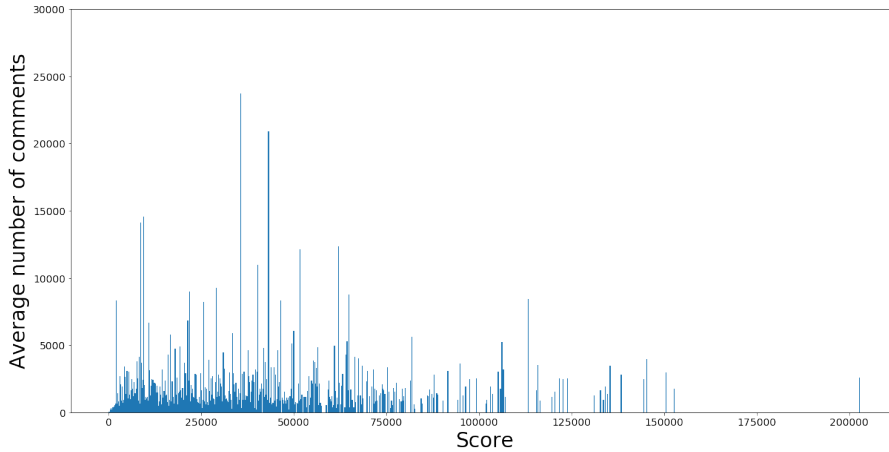

Figure 5: Distribution of comments to SFW posts against scores - Datasets regarding January and February 2019

Also in this case, we computed some parameters for the two distributions. They are shown in Table 10.

Parameter	SFW posts Jan-Feb	NSFW posts Jan-Feb	SFW posts Mar-Apr	NSFW posts Mar-Apr
Average score	9,881	4,191	8,809	3,819
Score of the last comment of the first quartile	2,035	1,157	1,993	1,215
Score of the last comment of the second quartile	4,686	2,357	4,551	2,484
Score of the last comment of the third quartile	11,106	4,486	9,953	4,667
Score of the last comment of the fourth quartile	202,696	69,591	209,154	71,566

Table 10: Parameters of the distributions of comments to posts against scores

A Wilcoxon rank sum test indicated that the score of comments for Jan-Feb SFW posts was statistically significantly higher than the corresponding one of NSFW posts ($\tau = 5.88 \cdot 10^{-5}, p < 0.01$).

The motivations behind this result are the same as those behind the knowledge extraction in Section 5.3.
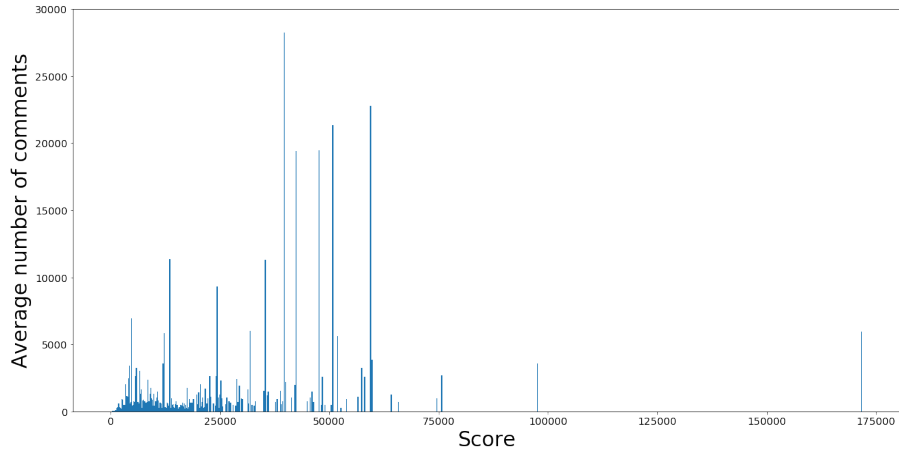
Figure 6: Distribution of comments to NSFW posts against scores - Datasets regarding January and February 2019

6.4 A deeper analysis of the stability of results

All the distributions we have seen so far are based on a data sample recovered from January 1st, 2019 to September 1st, 2019. Due to computational complexity reasons, we could not process the whole sample at the same time and, therefore, we divided it into bi-months, i.e. Jan-Feb and Mar-Apr. In all the distributions we have presented so far, we could verify that the Jan-Feb and Mar-Apr data led to very similar results. This is a strong remark of the stability of the results of our investigations.

However, before continuing with the next analyses, which will have an even higher computational complexity, we decided to carry out a further stability check. To this end, we considered all the posts published in Reddit from January 1st, 2019 to December 31st, 2019, and split them months by months. Then, for each month, we computed several parameters previously seen for the two bi-months. The results obtained are shown in Table 11 for SFW posts, and in Table 12 for NSFW posts. The analysis of these tables fully confirms that the results of our investigations are stable.

7 Investigating co-posting activity of the authors of NSFW posts

The goal of this analysis is to verify whether there is any correlation between the authors of NSFW posts. As usual, we will extract the information of interest and we will compare the behavior of authors of NSFW posts with the ones of SFW posts. In this activity, we will use a support data structure that we call *co-posting network*. Having observed in all the previous experiments that the results obtained for the Jan-Feb datasets (i.e., $\mathcal{D}$ and $\overline{\mathcal{D}}$) are stable, from now on we will refer to these two datasets only, avoiding to report the analysis of Mar-Apr datasets, too. In addition, since most of the operations that we will perform on the co-posting network are computationally expensive, we randomly extracted a subset $\mathcal{D}^*$ (resp., $\overline{\mathcal{D}}^*$) of $\mathcal{D}$ (resp., $\overline{\mathcal{D}}$) consisting of 75,000 SFW (resp., NSFW) posts to work on.

As a first task of this analysis, we give a formal definition of the co-posting network $\mathcal{P}$ (resp., $\overline{\mathcal{P}}$) built from the authors of SFW (resp., NSFW) posts stored in $\mathcal{D}^*$ (resp., $\overline{\mathcal{D}}^*$).

15

Parameter	Jan	Feb	Mar	Apr	May	Jun
GENERAL CHARACTERISTICS						
Number of authors who published at least one SFW post	391,898	387,458	365,785	389,154	387,562	374,531
Number of authors who published only SFW posts	380,261	374,564	359,851	378,582	377,423	365,751
Percentage of authors publishing SFW posts who published only posts of this type	97.03%	96.67%	98.37%	97.28%	97.38%	97.65%
Number of subreddits containing at least one SFW post	58,843	57,965	58,786	57,653	58,426	57,953
Number of subreddits containing only SFW posts	54,189	53,482	53,952	54,236	54,873	52,432
Percentage of subreddits containing SFW posts that contain only posts of this type	92.09%	92.22%	91.77%	94.07%	93.91%	90.47%
DISTRIBUTION OF SUBREDDITS AGAINST POSTS						
Maximum number of subreddits	47,480	47,116	47,996	49,502	48,294	47,733
Maximum number of posts	25,006	23,746	26,055	26,650	28,743	24,211
Average number of subreddits	125	120	154	141	133	118
Average number of posts	762	599	768	698	747	703
α (power law parameter)	1.6321	1.5806	1.7512	1.8358	1.6293	1.7024
δ (power law parameter)	0.0256	0.0238	0.0362	0.0357	0.0263	0.029
DISTRIBUTION OF AUTHORS AGAINST POSTS						
Maximum number of authors	555,854	559,602	566,139	540,511	551,863	541,585
Maximum number of posts	18,724	17,401	18,268	16,513	17,226	19,949
Average number of authors	2,106	1,862	2,280	2,164	2,021	2,209
Average number of posts	487	533	434	548	620	462
α (power law parameter)	1.4531	1.6718	1.3565	1.399	1.5478	1.3742
δ (power law parameter)	0.0465	0.0359	0.0545	0.0233	0.0428	0.0757
DISTRIBUTION OF POSTS AGAINST SCORES						
Maximum score	183,453	185,056	180,553	191,864	180,578	179,099
Average score	9,826	8,651	9,594	9,576	8,901	9,415
α (power law parameter)	1.5986	1.631	1.4672	1.6026	1.6507	1.5681
δ (power law parameter)	0.0189	0.0186	0.0198	0.0086	0.0179	0.0359
DISTRIBUTION OF SUBREDDITS AGAINST AUTHORS						
Maximum number of subreddits	62,839	65,934	70,585	65,861	63,087	62,325
Average number of subreddits	149	145	154	150	133	148
Maximum number of authors	20,285	19,571	18,808	21,801	20,029	19,801
Average number of authors	603	623	584	678	587	650
α (power law parameter)	1.7185	1.7064	1.6209	1.608	1.7013	1.7853
δ (power law parameter)	0.0298	0.0485	0.0315	0.02	0.0379	0.0327

Parameter	Jul	Ago	Sep	Oct	Nov	Dec
GENERAL CHARACTERISTICS						
Number of authors who published at least one SFW post	59,465	60,563	59,489	59,873	58,985	60,236
Number of authors who published only SFW posts	58,801	59,423	58,965	58,742	58,632	59,542
Percentage of authors publishing SFW posts who published only posts of this type	98.88%	98.11%	99.11%	98.11%	99.40%	98.84%
Number of subreddits containing at least one SFW post	89,360	87,953	89,236	88,462	87,932	88,167
Number of subreddits containing only SFW posts	82,050	82,587	85,496	83,647	83,146	84,963
Percentage of subreddits containing SFW posts that contain only posts of this type	91.82%	90.74%	93.68%	91.76%	91.7%	94.4%
DISTRIBUTION OF SUBREDDITS AGAINST POSTS						
Maximum number of subreddits	46,283	46,882	48,777	47,676	48,886	47,070
Maximum number of posts	22,261	19,071	23,642	29,330	26,346	28,419
Average number of subreddits	158	99	116	109	110	120
Average number of posts	794	889	814	704	748	713
α (power law parameter)	1.582	1.8481	1.7838	1.7313	1.5937	1.5125
δ (power law parameter)	0.0186	0.0305	0.0535	0.0329	0.0468	0.0154
DISTRIBUTION OF AUTHORS AGAINST POSTS						
Maximum number of authors	541,585	574,678	542,568	569,611	576,835	556,736
Maximum number of posts	16,823	19,320	18,692	18,460	16,499	17,766
Average number of authors	2,377	2,298	1,919	1984	2,008	2,123
Average number of posts	441	579	429	614	264	551
α (power law parameter)	1.3323	1.406	1.4688	1.4054	1.3093	1.525
δ (power law parameter)	0.0713	0.0491	0.0561	0.0424	0.064	0.038
DISTRIBUTION OF POSTS AGAINST SCORES						
Maximum score	194,305	176,975	164,394	186,004	172,001	177,739
Average score	10,449	9,926	9,103	9,813	8,434	9,345
α (power law parameter)	1.5089	1.5785	1.4772	1.6389	1.4331	1.6354
δ (power law parameter)	0.0114	0.054	0.0245	0.0389	0.0226	0.0012
DISTRIBUTION OF SUBREDDITS AGAINST AUTHORS						
Maximum number of subreddits	59,963	57,573	59,898	52,885	62,111	63,232
Average number of subreddits	145	144	163	155	153	154
Maximum number of authors	18,901	20,056	20,285	19,962	21,078	20,909
Average number of authors	686	673	811	543	611	651
α (power law parameter)	1.7622	1.6287	1.4544	1.8174	1.5256	1.7388
δ (power law parameter)	0.0159	0.0263	0.043	0.0254	0.0184	0.0378

Table 11: Monthly trend of some parameters related to SFW posts

Formally speaking,

Parameter	Jan	Feb	Mar	Apr	May	Jun
GENERAL CHARACTERISTICS						
Number of authors who published at least one NSFW post	36,758	35,452	36,542	36,874	36,863	36,453
Number of authors who published only NSFW posts	36,094	35,259	36,501	36,165	36,135	36,023
Percentage of authors publishing NSFW posts who published only posts of this type	98.19%	99.45%	99.88%	98.07%	98.02%	98.82%
Number of subreddits containing at least one NSFW post	41,365	40,985	41,298	41,547	41,235	40,958
Number of subreddits containing only NSFW posts	34,055	33,254	34,587	32,982	33,563	34,159
Percentage of subreddits containing NSFW posts that contain only posts of this type	82.33%	81.13%	83.74%	79.38%	81.39%	83.40%
DISTRIBUTION OF SUBREDDITS AGAINST POSTS						
Maximum number of subreddits	18,332	17,985	19,547	21,034	20,135	20,235
Maximum number of posts	34,424	32,547	31,854	31,329	30,896	32,541
Average number of subreddits	53	52	50	54	55	52
Average number of posts	892	890	896	901	895	890
α (power law parameter)	1.6896	1.6721	1.6874	1.6852	1.6796	1.6852
δ (power law parameter)	0.0258	0.0254	0.0251	0.0254	0.0214	0.0261
DISTRIBUTION OF AUTHORS AGAINST POSTS						
Maximum number of authors	131,070	130,152	131,250	133,594	131,452	132,654
Maximum number of posts	16,383	16,125	14,214	15,674	16,540	14,210
Average number of authors	437	432	435	441	432	436
Average number of posts	541	542	540	542	544	539
α (power law parameter)	1.5463	1.7985	1.6222	1.8407	1.9456	1.4833
δ (power law parameter)	0.03345	0.0233	0.0239	0.0639	0.0388	0.0458
DISTRIBUTION OF POSTS AGAINST SCORES						
Maximum score	106,947	146,561	75,657	112,830	105,566	66,095
Average score	8,805	9,170	7,123	7,885	9,287	10,197
α (power law parameter)	1.6062	1.5162	1.6933	1.8989	1.6951	1.4956
δ (power law parameter)	0.0145	0.0265	0.042	0.0611	0.0346	0.0139
DISTRIBUTION OF SUBREDDITS AGAINST AUTHORS						
Maximum number of subreddits	62,839	63,382	61,204	33,963	50,609	53,781
Average number of subreddits	150	151	140	148	162	163
Maximum number of authors	20,285	17,549	19,347	11,326	18,495	19,324
Average number of authors	603	600	636	533	538	647
α (power law parameter)	1.7156	1.7682	1.6166	1.9204	1.753	1.6321
δ (power law parameter)	0.0312	0.0241	0.0384	0.0236	0.0187	0.0418

Parameter	Jul	Ago	Sep	Oct	Nov	Dec
GENERAL CHARACTERISTICS						
Number of authors who published at least one NSFW post	37,165	35,986	36,432	36,540	36,354	36,589
Number of authors who published only NSFW posts	36,984	35,421	35,962	35,986	35,756	35,852
Percentage of authors publishing NSFW posts who published only posts of this type	99.51%	98.42%	98.77%	98.48%	98.35%	97.98%
Number of subreddits containing at least one NSFW post	41,542	40,986	41,246	41,258	40,983	41,496
Number of subreddits containing only NSFW posts	34,478	33,352	34,254	34,165	33,241	33,986
Percentage of subreddits containing NSFW posts that contain only posts of this type	82.99%	81.37%	83.04%	82.80%	81.10%	81.90%
DISTRIBUTION OF SUBREDDITS AGAINST POSTS						
Maximum number of subreddits	20,135	18,564	17,423	19,631	18,328	20,124
Maximum number of posts	30,451	32,598	30,125	29,874	34,210	32,021
Average number of subreddits	50	59	52	53	51	50
Average number of posts	891	885	891	889	893	891
α (power law parameter)	1.6236	1.6454	1.59874	1.6598	1.6432	1.6953
δ (power law parameter)	0.0265	0.0259	0.0298	0.0265	0.0264	0.0254
DISTRIBUTION OF AUTHORS AGAINST POSTS						
Maximum number of authors	130,254	134,250	133,247	132,478	136,587	131,489
Maximum number of posts	16,125	14,256	15,879	16,325	14,369	16,362
Average number of authors	436	435	431	442	429	432
Average number of posts	543	540	539	551	543	544
α (power law parameter)	1.6992	1.4551	1.5295	1.5527	1.5524	1.6091
δ (power law parameter)	0.0446	0.048	0.0201	0.0268	0.0031	0.0428
DISTRIBUTION OF POSTS AGAINST SCORES						
Maximum score	97,462	143,430	102,590	100,844	104,027	81,167
Average score	7,866	8,613	8,801	11,050	7,148	8,012
α (power law parameter)	1.6422	1.5874	1.4948	1.7059	1.7936	1.3969
δ (power law parameter)	0.040	0.028	0.0386	0.0324	0.0184	0.0354
DISTRIBUTION OF SUBREDDITS AGAINST AUTHORS						
Maximum number of subreddits	49,210	76,791	64,241	54,351	50,864	34,037
Average number of subreddits	127	146	136	170	120	139
Maximum number of authors	17,425	20,605	23,952	20,608	18,613	16,594
Average number of authors	592	591	657	708	600	545
α (power law parameter)	1.7653	1.7342	1.5258	1.9738	1.6143	1.5882
δ (power law parameter)	0.0317	0.037	0.0204	0.0371	0.0207	0.0401

Table 12: Monthly trend of some parameters related to NSFW posts

$$\mathcal{P} = \langle N, E \rangle \qquad \overline{\mathcal{P}} = \langle \overline{N}, \overline{E} \rangle$$

Here, N (resp., $\overline{N}$) is the set of the nodes of $\mathcal{P}$ (resp., $\overline{\mathcal{P}}$). There is a node $n_i \in N$ (resp., $\overline{N}$) for each author a_i of SFW (resp., NSFW) posts of $\mathcal{D}^*$ (resp., $\overline{\mathcal{D}}^*$). There is an edge $(n_i, n_j, w_{ij}) \in E$ (resp., $\overline{E}$) if the authors a_i and a_j (associated with n_i and n_j, respectively) submitted at least one post in the same subreddit. w_{ij} is the number of subreddits having at least one SFW (resp., NSFW) post of a_i and, simultaneously, at least one SFW (resp., NSFW) post of a_j.

Then, we calculated some of the basic parameters of $\mathcal{P}$ and $\overline{\mathcal{P}}$; they are shown in Table 13. From the analysis of this table, we can deduce that:

- The number of co-posting authors of NSFW posts is smaller than the number of co-posting authors of SFW posts.

- The authors of NSFW posts are more interconnected with each other. This is shown by both the density of $\overline{\mathcal{P}}$ (which is about three times the one of $\mathcal{P}$) and the average degree of $\overline{\mathcal{P}}$ (which is much greater than twice the degree of $\mathcal{P}$). As we will see in the following, this can be explained considering that they are authors belonging to a niche context.

- The average clustering coefficient of $\overline{\mathcal{P}}$ is greater than the one of $\mathcal{P}$, but not as much as the density. This suggests that in $\overline{\mathcal{P}}$ fewer triads are closed than in $\mathcal{P}$. This implies that, probably, in $\overline{\mathcal{P}}$ there are more "bridge" authors than in $\mathcal{P}$. These authors tend to act as intermediaries between other authors who do not know each other. They could be expert authors who cooperate with many new authors initially unknown to each other.

Parameter	$\mathcal{P}$	$\overline{\mathcal{P}}$
Number of nodes	59,465	36,758
Number of edges	3,164,169	5,398,082
Density	0.001789	0.007990
Maximum Degree	2,593	3,670
Average Degree	106.42	293.70
Average Clustering Coefficient	0.7388	0.7755

Table 13: Basic parameters of the co-posting networks $\mathcal{P}$ and $\overline{\mathcal{P}}$

After this, we computed the distribution of the nodes of $\mathcal{P}$ and $\overline{\mathcal{P}}$ against their degree centrality. The results obtained are reported in Figures 7 and 8.

From the analysis of these figures we can see that both distributions follow a power law. We computed the corresponding values of α and δ and obtained that $\alpha = 2.2929$ and $\delta = 0.0470$ for $\mathcal{P}$ and $\alpha = 2.6811$ and $\delta = 0.0678$ for $\overline{\mathcal{P}}$. These values tell us that the two distributions are similar.

Furthermore, looking carefully at the distributions in Figures 7 and 8, it emerges another unexpected, extremely peculiar, feature. In fact, we can observe some spikes. Excluding that these spikes are noise, they could be caused by the fact that the networks $\mathcal{P}$ and $\overline{\mathcal{P}}$ are actually disconnected and each network consists of a set of connected components. We found extremely interesting to check if this hypothesis was true. Therefore, we carried out this analysis and verified that, actually, we were right. In fact, we found that $\mathcal{P}$ consists of 15,952 connected components. Of these, 11,514 are made up of a single node. The maximum connected component includes 21,364 nodes (equal to 35,92% of the network nodes) and 2,909,206 arcs (equal to 91.94% of the network arcs). The distribution of the connected components against their size (i.e., the number of nodes they include) follows a power
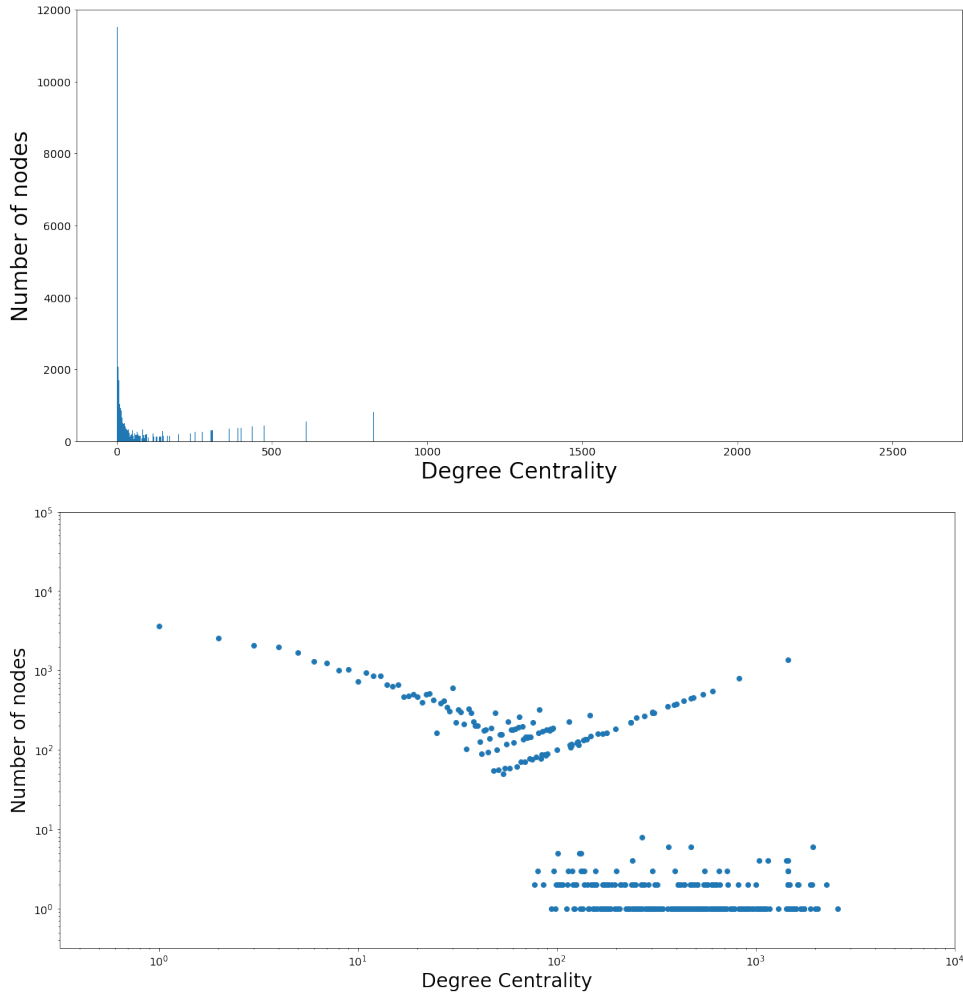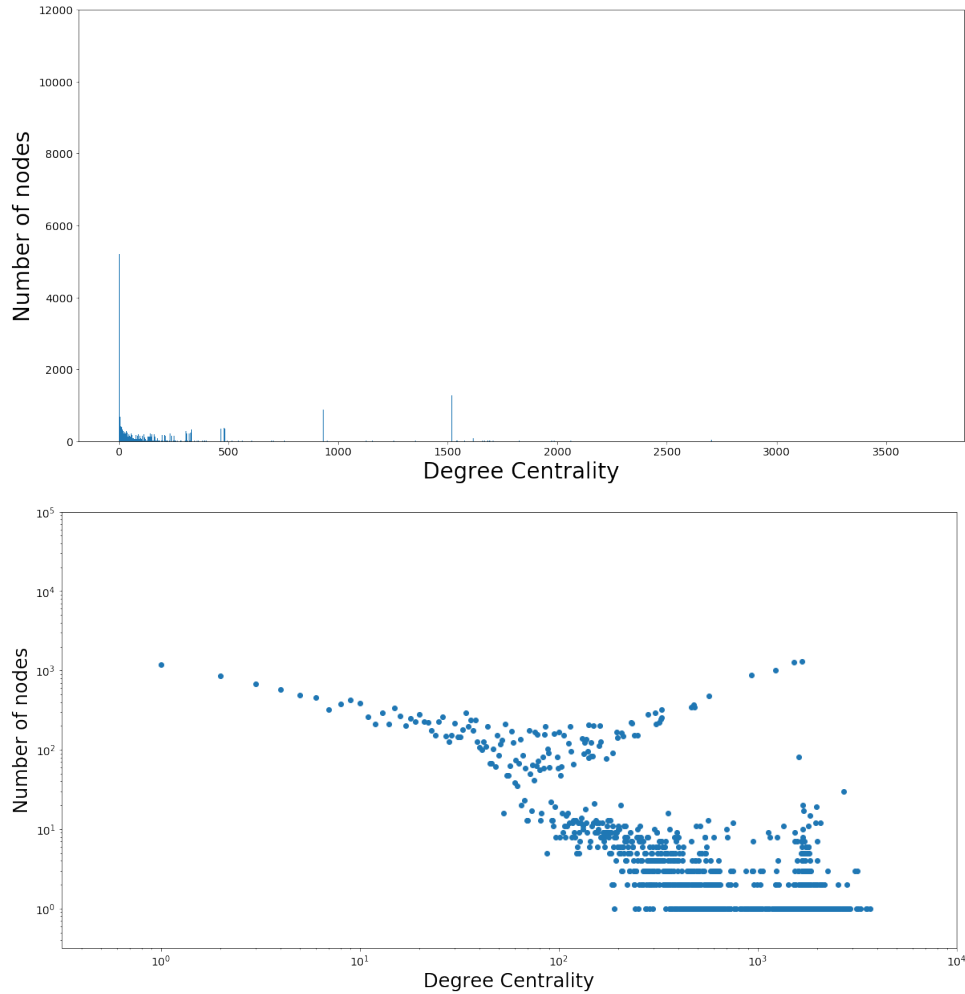
18

Figure 7: Distribution of the nodes of $\mathcal{P}$ against their degree centrality - linear scale (on top) and log-log scale (on bottom)

law with $\alpha = 1.562$ and $\delta = 0.060$. The network $\overline{\mathcal{P}}$ consists of 6,032 connected components, where 5,214 are made of a single node. The maximum connected component comprises 28,165 nodes (equal to 76.62% of the network's nodes) and 5,382,255 arcs (equal to 99.71% of the network's arcs). The distribution of the connected components against their size follows a power law with $\alpha = 1.548$ and $\delta = 0.065$.

The analysis of connected components strengthens some results obtained previously, in particular: *(i)* the number of co-posting authors of SFW posts is greater than the corresponding number of co-posting authors of NSFW posts; *(ii)* the authors of NSFW posts are more connected to each other (probably due to the presence of the "bridge" users mentioned above) than the ones of SFW posts.

At this point, we wanted to investigate more on the behavior of the authors of SFW and NSFW posts. Specifically, we treated three activities, namely the writing of posts, the tendency to publish on many subreddits and the ability to attract interest. For each of these activities, we selected the

Figure 8: Distribution of the nodes of $\overline{\mathcal{P}}$ against degree centrality - linear scale (on top) and log-log scale (on bottom)

top-ten authors from the maximum connected component of $\mathcal{P}$ and $\overline{\mathcal{P}}$ and we studied their behavior. In particular, Figure 9 (resp., 10 and 11) shows the top-ten authors who wrote the highest number of posts (resp., published in the largest number of subreddits, received the highest number of comments). The left part of this figure refers to the authors of SFW posts (belonging to the network $\mathcal{P}$), while the right part refers to the authors of NSFW posts (belonging to the network $\overline{\mathcal{P}}$).

These figures altogether outline a very precise author behavior. In fact, it can be noted that, regardless of the activity considered, the authors of SFW posts show a power law distribution, while the authors of NSFW posts show a very slowly decreasing distribution. This allows us to conclude that there are few very active authors of SFW posts and many inactive ones in Reddit. By contrast, there are many quite active authors of NSFW posts. Once again, it seems that these last tend to "team up" much more than the ones of SFW posts.

These results can be explained considering that the phenomenon of NSFW posts is a niche one
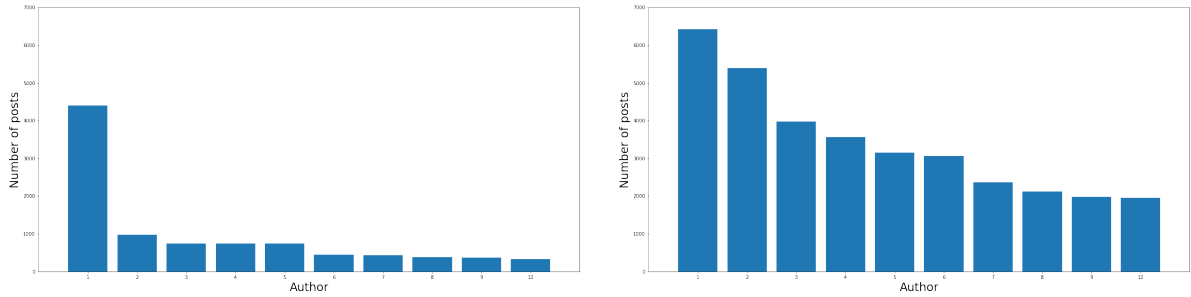
Figure 9: Top-ten authors who submitted more posts - authors of SFW posts at left and of NSFW posts at right
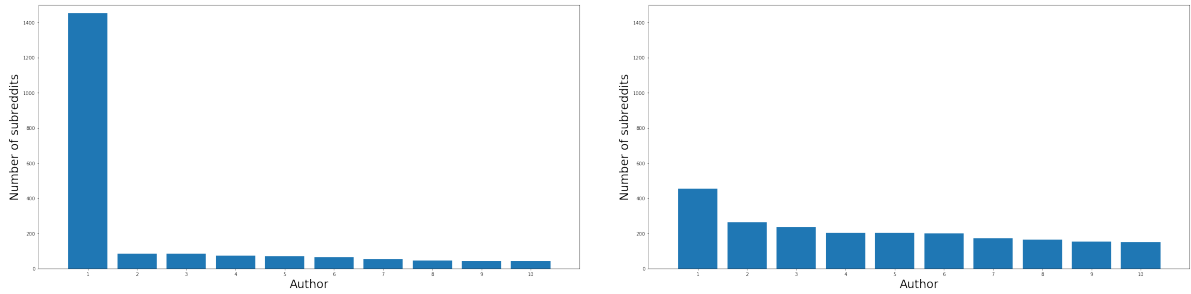

Figure 10: Top-ten authors who published on more subreddits - authors of SFW posts at left and of NSFW posts at right
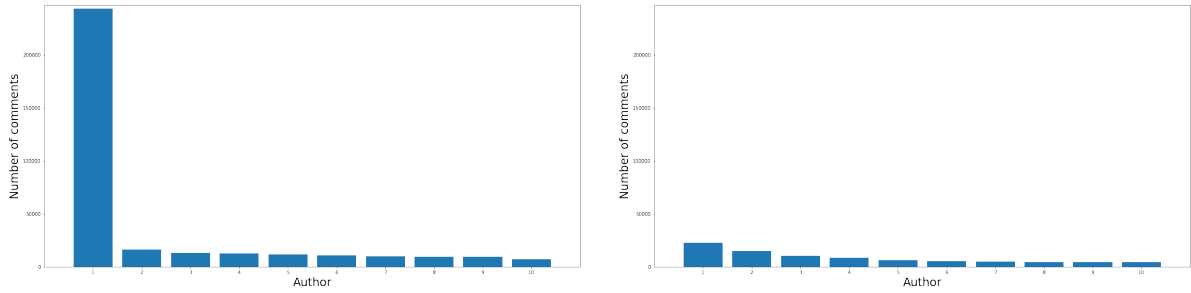

Figure 11: Top-ten authors who received more comments - authors of SFW posts at left and of NSFW posts at right

involving mostly particular kinds of user. These are very cohesive and form a fairly closed group. On the other hand, as we will see better in Section 9, all the knowledge extracted confirms this reasoning about the context behind NSFW posts.

8 Evaluating assortativity of the authors of NSFW posts

The concept of "assortativity", or "assortative mixing", in a social network points out the predilection of its nodes to be connected with other nodes that are somehow similar to them. This concept,

introduced by Newman [22], can be seen as an evolution of the concept of homophily [18], typical of Social Network Analysis. Assortativity is orthogonal to node similarity metrics considered, even if most of the authors in the literature have studied it with respect to node degree. According to this definition of assortativity, the nodes of a social network tend to be linked with other nodes having a degree similar to their own.

Assortativity is considered an extremely important property to be investigated by social network researchers. So we decided to analyze it for the authors of SFW and NSFW posts in Reddit. We would also pinpoint that: *(i)* like in the previous analyses performed in this paper, the goal is to characterize the assortativity of the authors of NSFW posts versus the one of the authors of SFW posts; *(ii)* the similarity property we decided to test for assortativity is node degree, because it is the most investigated one in the past literature on assortativity[3].

To carry out our assortativity analyses, we used the co-posting networks $\mathcal{P}$ and $\overline{\mathcal{P}}$ defined in Section 7. We showed the distributions of the nodes of these networks against degree centrality in Figures 7 and 8. As a first task, we sorted the authors of the two networks in descending order of degree centrality. After that, we splitted this ordered list into intervals. In particular, we considered 40 equi-width intervals $\{\mathcal{I}_1, \mathcal{I}_2, \cdots, \mathcal{I}_{40}\}$ for $\mathcal{P}$ and $\{\overline{\mathcal{I}}_1, \overline{\mathcal{I}}_2, \cdots, \overline{\mathcal{I}}_{40}\}$ for $\overline{\mathcal{P}}$. Since the number of nodes of $\mathcal{P}$ (resp., $\overline{\mathcal{P}}$) was 59,465 (resp., 36,578), each interval $\mathcal{I}_k$ (resp., $\overline{\mathcal{I}}_k$) contained 1,487 (resp., 915) authors[4].

At this point, we considered the interval $\mathcal{I}_1$ (resp., $\overline{\mathcal{I}}_1$) and, for each interval $\mathcal{I}_k$ (resp., $\overline{\mathcal{I}}_k$), we determined how many authors of $\mathcal{I}_1$ (resp., $\overline{\mathcal{I}}_1$) were connected to at least one author of $\mathcal{I}_k$ (resp., $\overline{\mathcal{I}}_k$). The results obtained are shown in Figure 12(a) (resp., 12(c)). Next, we computed the percentage of the authors of $\mathcal{I}_k$ (resp., $\overline{\mathcal{I}}_k$), who were connected to at least one author of $\mathcal{I}_1$ (resp., $\overline{\mathcal{I}}_1$). The results obtained are shown in Figure 12(e) (resp., 12(g)).

The analysis of Figures 12(a) and 12(e) shows a close correlation (i.e., a sort of backbone) between the authors of SFW posts with the highest degree centrality. On the contrary, the analysis of Figures 12(c) and 12(g) shows that this phenomenon does not occur for the authors of NSFW posts.

In order to evaluate the statistical significance of this result, we generated a null model to compare our outcomes with those of an unbiasedly random scenario. In particular, we built our null model shuffling the arcs of $\mathcal{P}$ (resp. $\overline{\mathcal{P}}$) among the nodes of this network. In this way, we left the original characteristics of $\mathcal{P}$ (resp. $\overline{\mathcal{P}}$) unchanged, except for the distribution of co-posting activities, which became unbiasedly random in the null model. The results obtained are shown in Figures 12(b), 12(d), 12(f) and 12(h).

Comparing Figures 12(b) and 12(f) with Figures 12(a) and 12(e) we can see that the represented distributions are similar. Indeed, many of the ranges with the highest values of Figures 12(a) and 12(e) continue to reach the highest values in Figures 12(b) and 12(f), too. However, these values are much smaller in the latter case. Therefore, we can conclude that the behavior observed in Figures 12(a) and 12(e) is not random, but intrinsic to $\mathcal{P}$ (and, therefore, to the authors of SFW posts in Reddit). On the contrary, if we consider Figures 12(c) and 12(g) (regarding the authors of NSFW posts in Reddit) and compare them with Figures 12(d) and 12(h), we can see that this phenomenon

---

[3]Actually, at the end of this section, for a further evidence of the results obtained, we also considered eigenvector centrality, beside degree centrality.

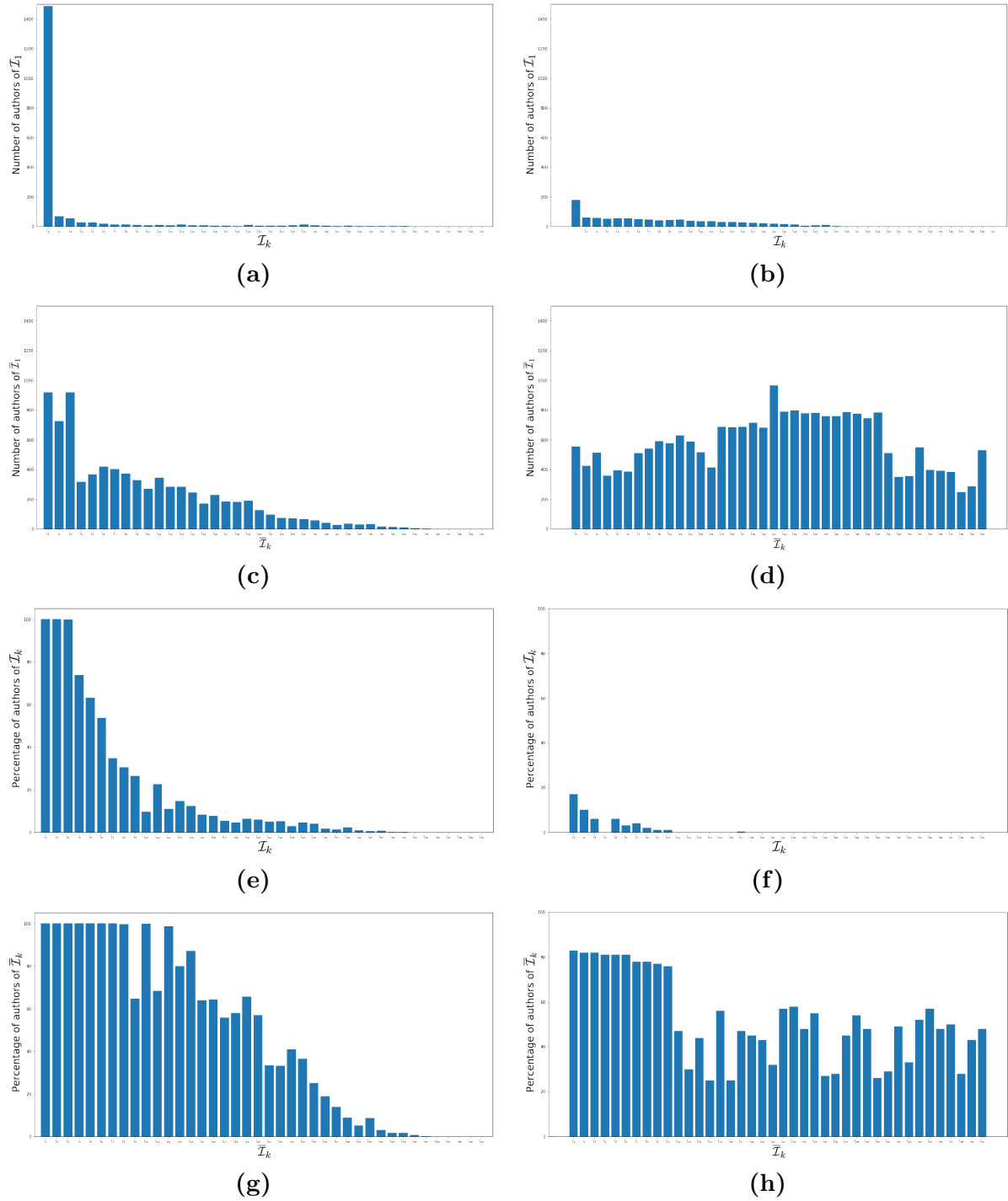[4]Actually, the last interval had a slightly smaller size equal to 1,472 (resp., 893) authors.

Figure 12: Degree Assortativity of the authors of NSFW and SFW posts (high degree authors)

does not occur for the authors of $\overline{\mathcal{P}}$.

The above analysis suggests that there is a degree assortativity among the authors of SFW posts

but not among the authors of NSFW posts. However, in order to confirm the assortativity of the authors of SFW posts, we need to verify whether this trend is still valid for the authors with an intermediate degree centrality and for those with a low degree centrality. If we want to make an exhaustive analysis, we should repeat the tasks previously performed for $\mathcal{I}_1$ (resp., $\overline{\mathcal{I}}_1$) for all the 40 intervals. For lack of space, we will limit our analysis to the intervals $\mathcal{I}_{20}$ (resp., $\overline{\mathcal{I}}_{20}$), as the representative of those with intermediate degree centrality, and $\mathcal{I}_{30}$ (resp., $\overline{\mathcal{I}}_{30}$), as the representative of those with low degree centrality[5].

Figure 13(a) (resp., 13(c)) shows the number of authors of $\mathcal{I}_{20}$ (resp., $\overline{\mathcal{I}}_{20}$) connected with at least one author of $\mathcal{I}_k$ (resp., $\overline{\mathcal{I}}_k$), while Figure 13(e) (resp., 13(g)) shows the percentage of authors of $\mathcal{I}_k$ (resp., $\overline{\mathcal{I}}_k$) connected with at least one author of $\mathcal{I}_{20}$ (resp., $\overline{\mathcal{I}}_{20}$). The analysis of these figures suggests the existence of a close correlation among the authors of SFW posts with an intermediate degree of centrality; this correlation does not exist for the authors of NSFW posts.

Even in this case, we compared these findings with those obtained in the null model. The latter are shown in Figures 13(b), 13(d), 13(f) and 13(h). Looking at all the diagrams reported in Figure 13, once again we can conclude that the observed behavior is not random, but it is a property of Reddit.

In the light of the last observation and of the previous conclusions on authors with an intermediate and a high degree centrality, we can certainly assert that there is no degree assortativity for the authors of NSFW posts. Instead, the possibility that such assortativity exists for the authors of SFW posts remains open.

In order to verify this last possibility, we carried out a study on the authors of $\mathcal{I}_{30}$. Figure 14(a) shows the number of authors of $\mathcal{I}_{30}$ connected to at least one author of $\mathcal{I}_k$, while Figure 14(c) shows the percentage of authors of $\mathcal{I}_k$ connected to at least one author of $\mathcal{I}_{30}$. These figures reveal the presence of a close correlation between the authors of SFW posts with a low degree centrality.

Even in this case, we compared the results obtained with those returned using the null model. We report the latter in Figures 14(b) and 14(d). The comparison of these figures with Figures 14(a) and 14(c) confirms that the behavior observed for these authors is an intrinsic property of Reddit.

Having verified that there is a sort of backbone among the authors of SFW posts with high (resp., medium, low) degree centrality, we can conclude that there is a degree assortativity for the authors of SFW posts in Reddit. Instead, this property is absent for the authors of NSFW posts in Reddit.

A further interesting analysis is to check if the tendency of the authors of SFW posts to be assortative and the tendency of the authors of NSFW posts to be not assortative is general or strongly depends on the type of assortativity that is being considered (in this case, degree assortativity).

As a premise to this discussion, it should be pointed out that every form of assortativity is independent, so it is impossible to come to a *general rule*. However, the analysis previously mentioned could surely lead us to discover some *trends*.

Therefore, we chose a second form of centrality (in particular, the eigenvector centrality) and we repeated all the steps previously taken for degree centrality with this second one.

The results obtained are very similar to those we have seen for degree centrality, i.e., we found the

---

[5]We did not choose the intervals $\mathcal{I}_k$ (resp., $\overline{\mathcal{I}}_k$), $k > 30$, because, during the analysis of the connected components, we saw that there is a high number of isolated nodes in $\mathcal{P}$ (resp., $\overline{\mathcal{P}}$) - see Section 7. Clearly, these nodes belong to the highest intervals and, if considered, could represent a bias in our analysis. To avoid this bias, we chose to not consider the intervals where they reside, and to select $\mathcal{I}_{30}$ (resp., $\overline{\mathcal{I}}_{30}$) as the representative of the intervals with low degree centrality.
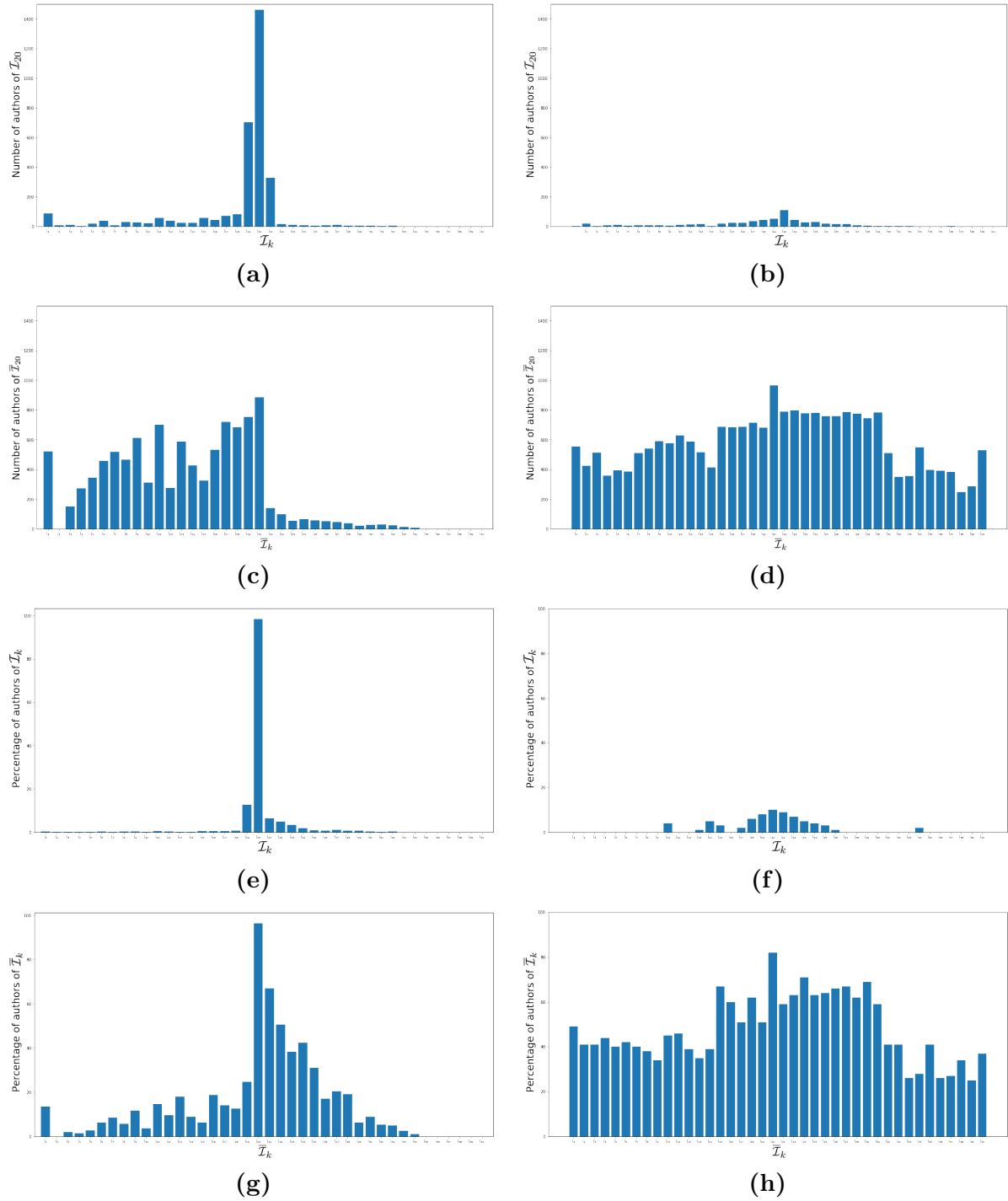
Figure 13: Degree Assortativity of the authors of NSFW and SFW posts (medium degree authors)

existence of a strong eigenvector assortativity for the authors of SFW posts and a lack of eigenvector assortativity for the authors of NSFW posts. For space reasons, we cannot show all the results.
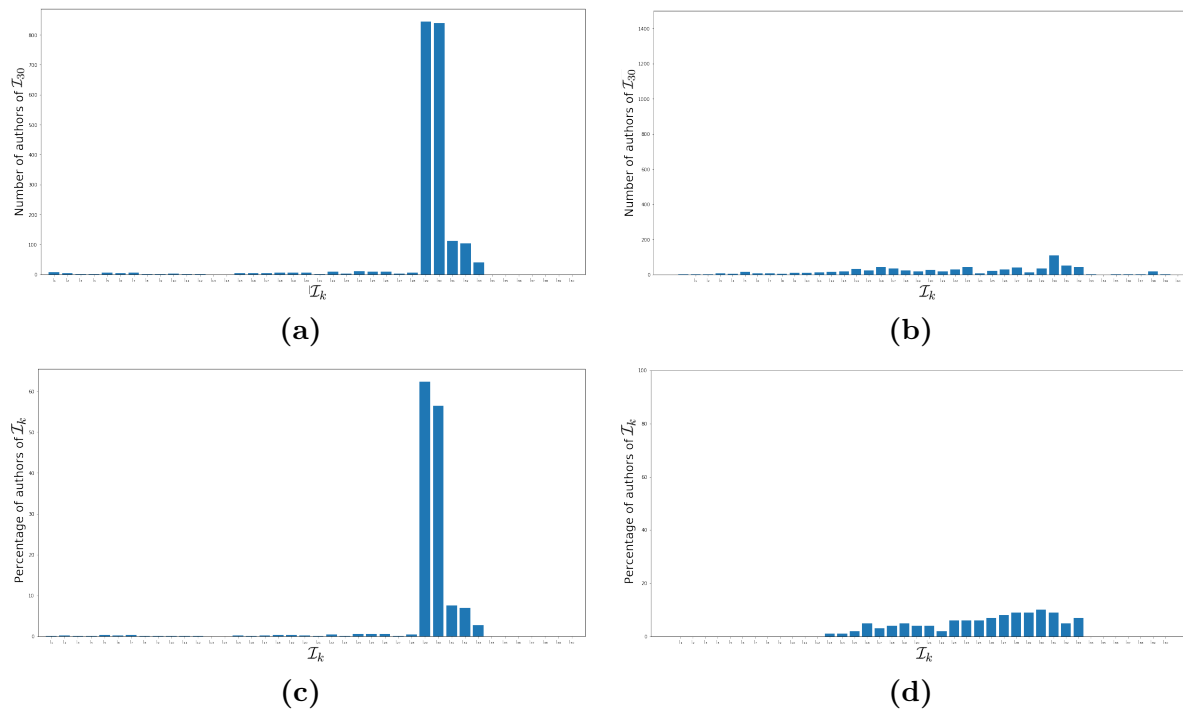
Figure 14: Degree Assortativity of the authors of SFW posts (low degree authors)

However, in order to give an idea of them, in Figure 15, we report what happens for authors with high eigenvector centrality. Comparing this figure with Figure 12, we can observe a strong similarity in the authors behavior in the two cases. As a consequence, we can say that SFW authors *tend* to be assortative, while NSFW authors *tend* to be not assortative.

This result can be explained by the strong community sense of the authors of NSFW posts. They are so cohesive that they do not feel the need to split into groups of peers. The most active people are still willing to interact with everyone else and not only with other equally active people.

9 Discussion

Combining together all the previous results, we can define three main findings related to posts, authors and subreddits, respectively. Some of these findings are made up of several sub-findings.

The three findings are the following:

PF (Finding on NSFW posts)

1. NSFW posts are generally published in much fewer subreddits, have much lower scores and are much less commented than SFW posts.

2. The scores of comments to NSFW posts are much lower than the ones to SFW posts.
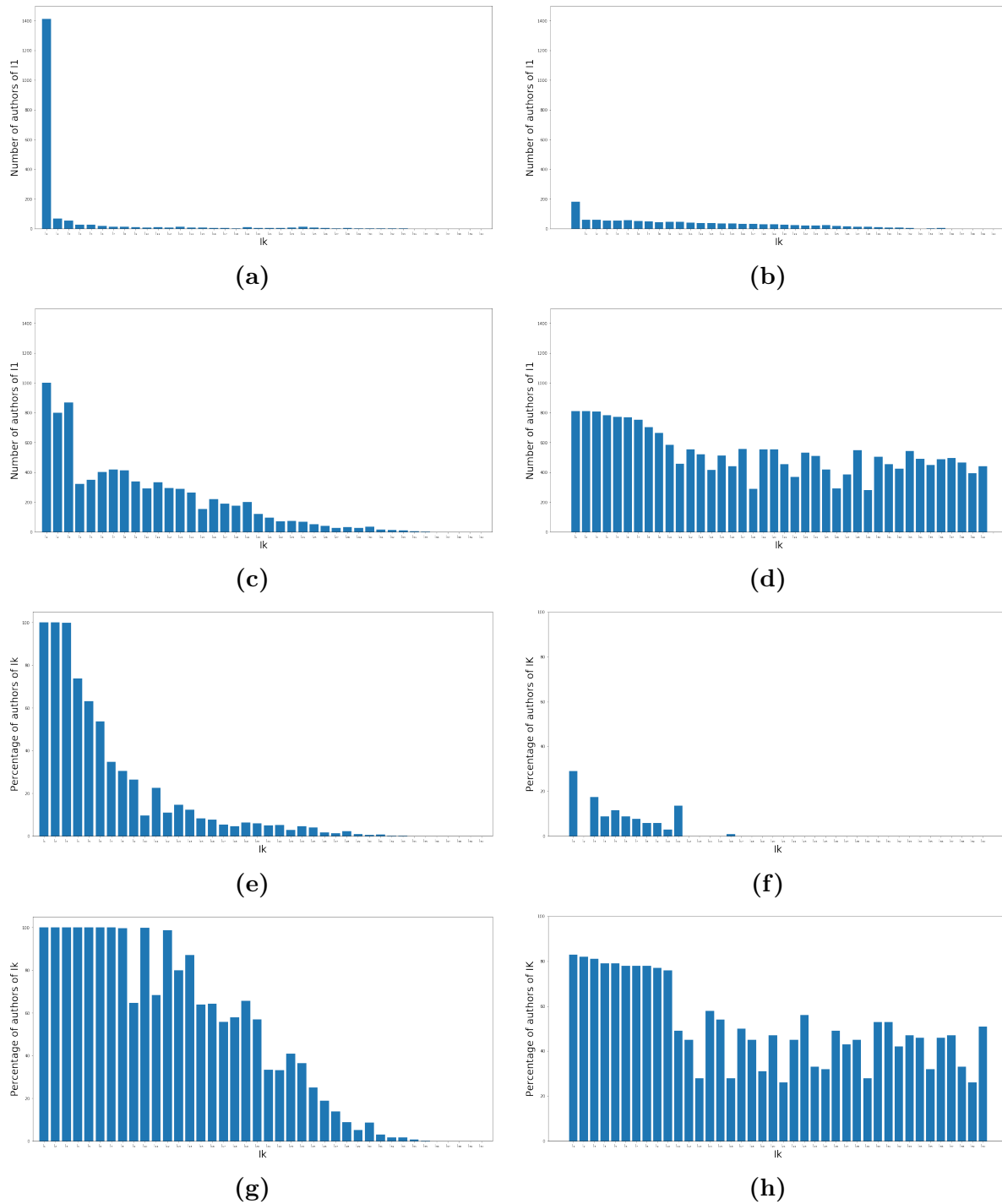
AF (Finding on NSFW authors)

Figure 15: Eigenvector Assortativity of the authors of NSFW and SFW posts (high degree authors)

1. NSFW authors tend: (i) to publish more posts, *(ii)* to publish in a fewer subreddits, *(iii)* to have a lower number of co-posting authors, *(iv)* to be more interconnected, active and

"teamed" than SFW authors.

2. The maximum number of negative posts published by a single NSFW author is much higher than the corresponding one of a single SFW author.

3. Differently from what happens to SFW authors, there is no degree assortativity and no eigenvector assortativity among NSFW authors.

SF (Finding on NSFW subreddits)

1. NSFW subreddits receive much fewer comments than SFW subreddits.

Now, we examine the previous findings in order to identify their correlations. This allows us to have a general view of the phenomenon of NSFW posts in Reddit.

The finding PF.1 tells us that an NSFW post is published in a limited number of subreddits. The finding AF.1 states that NSFW authors publish more than SFW ones. Now, since NSFW posts are fewer than SFW ones, we can conclude that NSFW posts have a much more limited number of authors. In addition, the combination of PF.1 and AF.1 is also a justification to the claim that NSFW authors publish in fewer subreddits than SFW authors.

Combining the findings PF.1 and AF.1 we can conclude that the phenomenon of NSFW posts is a niche one.

The finding PF.1 also tells us that the NSFW posts are little appreciated; actually, this information was quite expected. The results expressed by the finding PF.1 are reinforced by the finding AF.2, which tells us that the maximum number of negative posts published by a single NSFW author is greater than the corresponding number of an SFW author. The finding AF.2 is also, in part, a direct consequence of the finding AF.1.

The finding SF.1, stating that the NSFW subreddits receive fewer comments than SFW ones, represents a further confirmation of what the findings AF.1 and PF.1 say about the fact that NSFW posts are a niche phenomenon.

The poor consideration for NSFW posts, expressed by the finding PF.1, is further confirmed by the finding PF.2, which tells us that not only NSFW posts, but even comments to these posts, receive a much lower score than the comments to SFW posts.

The finding AF.1 (which tells us that the number of co-posting NSFW authors is fewer than SFW authors and that NSFW authors are more interconnected, active and "teamed" than SFW ones) represents a further confirmation that the NSFW post phenomenon is a niche one, carried out by few authors. However, it also tells us that these authors are very active and very well interconnected, ready to play "teamwork".

The last finding extracted, i.e., the finding AF.3, specifies that there is no degree or eigenvector assortativity for NSFW authors. In other words, the strong connection existing among NSFW authors is so widespread and compact that it does not let authors group into "narrow circles". In fact, the sense of cooperation between these authors is so high that the most active ones still collaborate with everyone else and do not limit their interactions to only those with their direct peers, as often happens in many other contexts.

10 Conclusion

In this paper, we have presented an approach to investigate NSFW posts in Reddit. We have seen that this type of content is frequent in this social medium and, despite this, there are very few studies on this subject in the past literature. We have tried to fill this gap and we have proposed an approach that investigates the phenomenon of NSFW posts in Reddit with descriptive, co-posting and assortativity analyses.

In this way, we have obtained three findings, which, together with the principles underlying our approach, are certainly the two main contributions of this paper. In fact, the findings reported in this paper provide valuable knowledge to better understand this phenomenon still little investigated. In addition, our way of proceeding defines a methodology that can be used to uncover the dynamics underlying NSFW contents in other social media.

In the future, there are several possible developments of our research efforts. First, it is possible to apply the proposed approach to other social media managing NSFW contents. In addition, we could extend our study of NSFW posts including an in-depth analysis of their content from a semantic point of view. Similarly, we could deepen our knowledge on the authors of NSFW posts applying sentiment analysis techniques to the posts they wrote or commented. Finally, we could consider to define a Machine Learning based approach to automatically identify and label NSFW posts, authors and communities, particularly when NSFW posts are not manually labeled by users. This last application can become extremely important to prevent NSFW contents from being sneakily and deceptively offered to unsuitable users (e.g., children).

References

[1] J. Baumgartner, S. Zannettou, B. Keegan, M. Squire, and J. Blackburn. The pushshift Reddit dataset. In *Proc. of the International AAAI Conference on Web and Social Media (ICWSM'20)*, volume 14, pages 830–839, Atlanta, GA, USA, 2020. AAAI Press.

[2] A.Q. Bhatti, M. Umer, S. H. Adil, M. Ebrahim, D. Nawaz, and F. Ahmed. Explicit Content Detection System: An Approach towards a Safe and Ethical Environment. *Applied Computational Intelligence and Soft Computing*, page 1463546, 2018. Hindawi.

[3] C. Buntain and J. Golbeck. Identifying Social Roles in Reddit Using Network Structure. In *Proc. of the International Conference on World Wide Web (WWW 2014)*, page 615–620, Seoul, Korea, 2014. ACM.

[4] M. Carpenter and M. Garner. NSFW: An Empirical Study of Scandalous Trademarks. *Cardozo Arts & Ent. LJ*, 33:321, 2015. HeinOnline.

[5] N. Cassavia, E. Masciari, C. Pulice, and D. Saccà. Discovering User Behavioral Features to Enhance Information Search on Big Data. *ACM Transactions on Interactive Intelligent Systems*, 7(2), 2017. ACM.

[6] T. Connie, M. Al-Shabi, and M. Goh. Smart content recognition from images using a mixture of convolutional neural networks. In *IT Convergence and Security 2017*, pages 11–18. 2018. Springer.

[7] D. Correa, L. A. Silva, M. Mondal, F. Benevenuto, and K. P. Gummadi. The many shades of anonymity: Characterizing anonymous social media content. In *Proc. of the International AAAI Conference on Web and Social Media (ICWSM 2015)*, pages 71–80, Oxford, UK, 2015. AAAI.

[8] S. Datta and E. Adar. Extracting Inter-Community Conflicts in Reddit. In *Proc. of the International Conference on Web and Social Media (ICWSM 2019)*, pages 146–157, Munich, Germany, 2019. AAAI.

[9] P.V.A. de Freitas, G.N.P. Santos, A.J.G. Busson, A.L.V. Guedes, and S. Colcher. A baseline for NSFW video detection in e-learning environments. In *Proc. of the Brazillian Symposium on Multimedia and the Web (WebMedia 2019)*, pages 357–360, Rio de Janeiro, Brazil, 2019. ACM.

[10] M. Fire and C. Guestrin. The rise and fall of network stars: Analyzing 2.5 million graphs to reveal how high-degree vertices emerge over time. *Information Processing & Management*, 57(2):102041, 2020. Elsevier.

[11] A. Grewal and J. Lin. The evolution of content analysis for personalized recommendations at Twitter. In *Proc. of the International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '18)*, pages 1355–1356, Ann Arbor, MI, USA, 2018. ACM.

[12] A. Guimaraes, O. Balalau, E. Terolli, and G. Weikum. Analyzing the Traits and Anomalies of Political Discussions on Reddit. In *Proc. of the International Conference on Web and Social Media (ICWSM 2019)*, pages 205–213, Munich, Germany, 2019. AAAI.

[13] Q. He, X. Wang, F. Mao, J. Lv, Y. Cai, M. Huang, and Q. Xu. CAOM: A community-based approach to tackle opinion maximization for social networks. *Information Sciences*, 513:252–269, 2020. Elsevier.

[14] Y. Kou, C.M. Gray, A.L. Toombs, and R.S. Adams. Understanding Social Roles in an Online Community of Volatile Practice: A Study of User Experience Practitioners on Reddit. *ACM Transactions on Social Computing*, 1(4):17:1–17:22, 2018. ACM.

[15] J. LaViolette and B. Hogan. Using Platform Signals for Distinguishing Discourses: The Case of Men's Rights and Men's Liberation on Reddit. In *Proc. of the International Conference on Web and Social Media (ICWSM 2019)*, pages 323–334, Munich, Germany, 2019. AAAI.

[16] Y. Li, Z. Su, J. Yang, and C. Gao. Exploiting similarities of user friendship networks across social networks for user identification. *Information Sciences*, 506:78–98, 2020. Elsevier.

[17] J.N. Matias. Going dark: Social factors in collective action against platform operators in the Reddit blackout. In *Proc. of the International Conference on Human Factors in Computing Systems (ACM CHI 2016)*, pages 1138–1151, San Jose, CA, USA, 2016. ACM.

[18] M. McPherson, L. Smith-Lovin, and J.M. Cook. Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27:415–444, 2001. JSTOR.

[19] A.N. Medvedev, R. Lambiotte, and J.C. Delvenne. The Anatomy of Reddit: An Overview of Academic Research. In *Dynamics On and Of Complex Networks III*, pages 183–204, Cham, 2019. Springer International Publishing.

[20] B. K. Narayanan and M. Nirmala. Adult content filtering: Restricting minor audience from accessing inappropriate Internet content. *Education and Information Technologies*, 23(6):2719–2735, 2018. Springer.

[21] E. Newell, D. Jurgens, H.M. Saleem, H. Vala, J. Sassine, C. Armstrong, and D. Ruths. User Migration in Online Social Networks: A Case Study on Reddit During a Period of Community Unrest. In *Proc. of the International Conference on Web and Social Media (ICWSM 2016)*, pages 279–288, Cologne, Germany, 2016. AAAI.

[22] M.E.J. Newman. Clustering and preferential attachment in growing networks. *Physical Review E*, 64(2):025102, 2001. APS.

[23] A. Nocera and D. Ursino. PHIS: a system for scouting potential hubs and for favoring their "growth" in a Social Internetworking Scenario. *Knowledge-Based Systems*, 36:288–299, 2012. Elsevier.

[24] Q. Shen and R. Carolyn. The Discourse of Online Content Moderation: Investigating Polarized User Responses to Changes in Reddit's Quarantine Policy. In *Proc. of the International Workshop on Abusive Language Online (ALW 2019)*, pages 58–69, Florence, Italy, 2019. Association for Computational Linguistics.

[25] P. Singer, F. Flöck, C. Meinhart, E. Zeitfogel, and M. Strohmaier. Evolution of Reddit: From the Front Page of the Internet to a Self-Referential Community? In *Proc. of the International Conference on World Wide Web (WWW 2014)*, page 517–522, Seoul, Korea, 2014. ACM.

[26] A. Soliman, J. Hafer, and F. Lemmerich. A Characterization of Political Communities on Reddit. In *Proc. of the ACM Conference on Hypertext and Social Media (HT'19)*, page 259–263, Hof, Germany, 2019. ACM.

[27] C. Tan and L. Lee. All Who Wander: On the Prevalence and Characteristics of Multi-Community Engagement. In *Proc. of the International Conference on World Wide Web (WWW 2015)*, page 1056–1066, Florence, Italy, 2015. ACM.

[28] K. Tiidenberg. Boundaries and conflict in a NSFW community on tumblr: The meanings and uses of selfies. *New Media & Society*, 18(8):1563–1578, 2016. Sage Publications.

[29] T. Weninger. An exploration of submissions and discussions in social news: mining collective intelligence of Reddit. *Social Network Analysis and Mining*, 4:173–192, 2014. Springer.

[30] F. Wilcoxon. Individual Comparisons by Ranking Methods. In *Breakthroughs in statistics*, pages 196–202. 1992. Springer.

[31] Y. Wu, H. Huang, N. Wu, Y. Wang, M.Z.A. Bhuiyan, and T. Wang. An incentive-based protection and recovery strategy for secure big data in social networks. *Information Sciences*, 508:79–91, 2020. Elsevier.

[32] D. Zhelonkin and N. Karpov. Training Effective Model for Real-Time Detection of NSFW Photos and Drawings. In *Proc. of the International Conference on Analysis of Images, Social Networks and Texts (AIST 2019)*, pages 301–312, Kazan, Russia, 2019. Springer.

[33] B. Zheng, O. Liu, J. Li, Y. Lin, C. Chang, B. Li, T. Chen, and H. Peng. Towards a distributed local-search approach for partitioning large-scale social networks. *Information Sciences*, 508:200–213, 2020. Elsevier.