

**Alex Bradley**

**Abstract**

Online forums can be a source of support for people with gambling-related problems. Forum threads contain detailed information about these gamblers' experiences. However, because of limitations in data collection and analysis, there have been few systematic analyses of forum content. The aim of this study is to use web scraping and correlated topic modelling to develop a bottom-up, data-driven approach to identify key issues raised by gamblers participating in an online forum, taking 2,298 posts from 1,400 unique authors over a twelve-year period. The data revealed ten themes that fall into four superordinate categories: negative emotions caused by gambling, the process of recovery, gambling products and money related concerns. Negative emotions associated with gambling was the most common topic occurring in 25% of posts. The process of recovery theme could be divided into formal and informal resources for dealing with gambling problems. Gambling products captured both traditional high street and new online forms of gambling. A final theme highlighted how family and friends become sources of finance to fund gambling. These findings can be used to design brief psychosocial education programs which highlight the consequences of gambling on oneself, one's family and the emotional impact that emerges from gambling.

Keywords: Gambling, Online Forums, Web Scraping, Correlated Topic Models

Abstract Word Count: 199

Total Word Count: 4,648

## **Introduction**

The majority of people experiencing gambling related harm or disordered gambling do not seek face-to-face treatment (Melville, Casey, & Kavanagh, 2007). It also appears unlikely gambling related harms will be incidentally picked up by healthcare practitioners, such as in primary care (Brown et al., 2016; Cowlshaw, Gale, Gregory, McCambridge, & Kessler, 2017). As such, for most people undergoing difficulties with gambling, they are likely to seek self-directed care (Gainsbury & Blaszczynski, 2011; Rodda, Lubman, & Dowling, 2017a). Computers and the internet have facilitated self-directed care and recovery through information websites and peer to peer support forums (Gainsbury & Blaszczynski 2011). The information held on peer-peer support forums is a vast source of potentially invaluable data but that also poses a challenge to researchers on how they can process and analyse large quantities of text (Adjerid & Kelley, 2018; Paxton & Griffiths, 2017). The aim of this paper is to use web scraping to determine the issues discussed by gamblers on an online support forum, using an unsupervised text processing method known as topic modelling. Topic modelling analyses the covariance between words in a text to determine the key issues raised within a wider body of text, in this case, posts on a gambling forum. This allows a data-driven approach to be used for the first time, to identify the key issues spontaneously raised by the forum users. Further, this approach complements the existing literature on gambling related harms, which uses a grounded theoretical approach (Langham et al., 2016). The rest of the introduction shall first review the current evidence on how gamblers use online forums and the potential benefits they have, and identify the challenges posed to researchers in making use of this wealth of rich detailed textual data.

### **Exploring interactions on support forums**

The previous literature analysing online forum content has attempted to categorise how users interact with each other. Content analyses of forum threads have revealed that most interactions on two online gambling support forums evolved around giving information and advice, being supportive to one another, asking for help, and sharing personal stories (Wood & Wood, 2009). Analysis of online posts has also found that gamblers talk about and discuss a wide range of range of change strategies to help control their gambling behaviour (Rodda et al., 2018). These change strategies were

found to fall across a continuum from pre-decisional strategies like recognising external and internal barriers to change to actional strategies like avoiding gambling stimuli in their environment or behavioural substitution (Rodda et al., 2018). Discursive analysis of postings to an online forum (Mudry & Strong, 2013) reported six themes within the online interactions: feelings of shame and guilt for their gambling behaviour and how it had affected the lives of others, conveying the causes of their problem gambling, reflecting on the recreational vs problematic aspects of gambling, discussing gambling as an addiction, control and responsibility over one's gambling and talking about recovery as an ongoing process. Focusing specifically on the emotional rather than the functional or experiential aspects discussed on gambling forums, Andrea (2015) presented an analysis of 24 stories collected from Gambler's Anonymous posts in Italy. Five clusters emerged: guilt resulting from the consequences of gambling, obsessions with gambling, perceiving gambling as a problem that needs to be addressed, viewing gambling as a risky behaviour and the intersection between gambling and life events.

This literature suggests there are some commonalities in the themes raised on online forums, in particular with the discussion of guilt and the discussion of problematic gambling. However, these are based on in-depth analyses of a small number of posts, which may not capture the totality of activity on a forum, and are a potential reason for the general lack of overlap between the different purposes (i.e. how people interact vs. the emotional content they use) in addition to their divergent analytical approaches (i.e. content vs discourse analyses). This highlights the need for a more comprehensive analysis of the content of online support forum posts.

### **Benefits of using online support forums**

The existing evidence suggests that users report the forums to be beneficial. Analysis of users of GAweb, an online peer-peer support forum based on Gamblers Anonymous, found their interactions created a new sense of hope for the future, an increased likelihood of seeking further face-face help, formed new friendships and a desire not to relapse so as not to let down the community (Cooper, 2004). It was further noted that those lurking (forum readers who do not post) benefitted from reading the social interactions on the forum through an increased likelihood of disclosing gambling problems

to others in the future (Cooper, 2004). The communal aspect of online forums also appears to help users feel less alone with their gambling problems, and enable them to develop insights into their own behaviour (Wood & Wood, 2009). There is also evidence to suggest that using online forums in conjunction with other self-directed help (i.e. chat, emails, self-help and information websites) resulted in the reporting of less gambling symptoms, reduced frequency of gambling and spending less on gambling (Rodda, Lubman, & Dowling, 2017). These findings highlight how online forums can be a beneficial source of support for gamblers, as forum engagement fosters the growth of relationships that appear to aid attempts to control or cease gambling. Thus identifying the themes discussed on online forums is likely to be informative for tailoring efficacious brief interventions and education programs.

### **Challenges to researchers using online peer-peer forums**

Although forums are commonly used, and contain a rich pool of data from gamblers across the world who may be marginalised in the gambling literature, there are difficulties in using them for research. Textual data can be difficult for researchers to statistically model, and the quantity of available data can make getting a holistic picture of the whole using conventional qualitative approaches challenging. Conventional qualitative techniques typically select small subsets of the data for analysis. We therefore outline and introduce how unsupervised methods of text processing known as topic modelling can be used to explore the large quantities of rich detailed data now available to researchers on the web. This aims to encompass the breadth of data that is characteristic of quantitative research, whilst retaining some of the contextual depth that qualitative research offers. In doing so, we aim to enumerate the different topics that emerge from a large scale analysis of posts from on an online forum.

Although this is primarily an exploratory analysis, the literature suggests that a number of themes are likely to emerge in this analysis. What is common across these studies is that negative emotions from gambling (especially shame and guilt), the effect gambling has on the gambler's life and particularly their family, and providing sources of support are likely to emerge as topics from this analysis as well.

## Method

### Data Source

The sample is comprised of the first posts from 2,298 threads on the 'My Journal' forum from a UK online gambling help website called Gambling Therapy. The 'My Journal' forum is described as a place where people can post about their life before, during and after experiencing gambling problems. All of the threads were scraped from the first forum post (30<sup>th</sup> June 2005) to the date of web scraping (16<sup>th</sup> November 2017). Four threads pinned to the top of the forum were removed since their content was provided by moderators not users of the site (2,294 threads). The average length of a thread was 308.4 words ( $SD = 333.55$ ). Over the 2,294 threads, there were 1,400 unique authors and the average number of posts per author was 1.64 ( $SD = 2.55$ ). Ethics approval for this project was received from Nottingham University Psychology Ethics Committee (Ref: F969).

### Data analysis

#### Web scraping procedure

The 92 pages of the My Journal Forum on GamblingTherapy.org were web scraped (i.e. extracted from the website) on the 16<sup>th</sup> November 2017. On the first 91 pages, there were 25 threads and 23 threads on the last page. The web scraper extracted from each of the pages the Title and Number of Replies. On each of the 92 pages, the web scraper followed the link in the thread title to the first page of each thread and extracted the author of the threads (i.e. username), date posted and text of the original post that started the thread. To protect the anonymity of the forum users the author (username), date of post and title of the post were subsequently deleted after data processing. Only the original post of each thread was retained and used in the statistical analysis.

All web scraping was done using Rstudio and the 'rvest' package (RStudio Team, 2015; Wickham, 2015). The web scraping script, and analysis code is available on the Open Science Framework (<https://osf.io/6ymqg/>) in the folder entitled 'Real Life Example of Web Scraping a Gambling Forum'.

**Pre-processing of the text.**

The web scraping produces a corpus of text, a collection of the words from each of the initial posts. However, this raw data requires cleaning before analysis. Before performing topic modelling we performed common text processing techniques like removing numbers, punctuation (i.e. full stops, question marks, exclamation marks but not apostrophes since they can alter the meaning of words for example can't with the apostrophe omitted becomes cant), converting the letters to lower case, stemming the words (reducing words to their base form or root i.e. "gambling" and "gambler" become "gambl"), removing stop words (i.e. "the" and other articles, "and" and other conjunctions, and other words that occur commonly but infer little meaning), removing words with less than three characters (numbers, letters or special punctuation like apostrophes) and removing sparse terms (defined as those that occur in less than 10 documents) (Grimmer & Stewart, 2013). The corpus was then transformed into a Document-Term Matrix (a matrix where a column represents a document, a row represents a word and a cell represents the frequency of the word in that document) ready for analysis. The final corpus contained 2,260 documents with 16, 257 unique words.

**Overview of the Analysis**

The text from the threads was analysed using a correlated topic model (Blei & Lafferty, 2007). Topic modelling is designed to identify multiple categorical, unobserved (or 'latent') populations within an observed dataset, a type of statistical analysis called mixture modelling. Topic models take multiple pieces of text, calculate the co-occurrence of words within them, and segment the vocabulary of words into a number of distributions (or topics) (Blei et al., 2003; Grün & Hornik, 2011). These are groupings of words that appear together in the texts entered into the analysis with a common underlying meaning (Grimmer & Stewart, 2013). For example, the models might identify a topic related to sports betting. In such a topic, some words might appear together frequently ('bet', 'football', 'punt'), whereas there might be words related to gambling that appear less with them (e.g. 'slot', 'chance', 'pokie'). Topic modelling is a 'mixed membership' model, because the words entered into the model are more or less representative of every topic, rather than being assigned to a topic in a mutually exclusive fashion (Roberts et al., 2014). Similar to other mixture models, correlated topic

models may fail to converge or reach local solutions, a scenario called ‘multimodality’. The correlated topic model was performed using spectral initialization to mitigate issues of multimodality from choosing sub-optimum starting values (Roberts, Stewart, & Tingley, 2015). One of the risks with the estimation of complex models using maximum likelihood estimation is that identifies a local rather than a global estimate. Spectral initialization is a way of choosing starting values for model estimation that will yield stable results not determined by the initial starting values. In addition to identifying and interpreting the topics, which are the main results of such an analysis, the most frequent words in each topic that are mutually exclusive to other topics (or FREX) can be used to illustrate the meaning of the identified topics. The output of the topic model was used also to identify the most representative post to each topic. The topic modelling was conducted using the ‘stm’ package in R (Roberts, Stewart, & Tingley, 2015.).

### **Deciding on the number of topics**

There is no right or wrong answer to how many topics are best (Grimmer & Stewart, 2013), and models will differ on indices of fit (e.g. better held out likelihood, lower residuals, higher semantic coherence, and higher exclusivity) (Airoldi & Bischof, 2016; Mimno, Wallach, Talley, Leenders, & Mccallum, 2011; Roberts et al., 2014). These measures of fit are similar to those used in other latent variable analyses (e.g. factor analysis). Held out likelihood is a measure of cross-validation – the analysis is done on a subset of the data to examine the robustness of the models obtained. Residuals are a direct measure of model fit (smaller residuals, less error). Coherence is a metric of the semantic relations within a topic, which is a metric of the quality of the topics generated by the model (Mimno et al., 2011). However, it has been noted that high semantic coherence can be attained with a small number of topics with commonly used words (Roberts et al., 2014). Exclusivity refers to the extent to which the most representative words are unique to a topic. Models that have greater exclusivity are preferred because they provide non-redundant information. Because words are more or less representative of every topic, this assesses the extent to which the topics are orthogonal. Roberts et al (2014) advise that a combination of exclusivity and semantic coherence is used to judge the best number of topics to retain. We used the ‘searchK’ function within the stm package in R to identify the

most appropriate number of topics.

[Insert Figures 1, and 2]

## **Results**

A ten-topic model had a reasonable held out likelihood, the lowest residual score and offered the highest exclusivity and reasonable semantic coherence (see Figure 1 and Figure 2). These indices of fit each show that a ten topic model is the best fit of the data and is more readily interpretable than other alternatives. Table 1 shows the top twenty words with the highest FREX scores over the ten topics. Figure 3 shows the highest FREX words over the topics and the expected proportions of the topics in the documents included in the model. To streamline interpretation of the findings, we then grouped them into five, overarching themes: negative emotions caused by gambling, resources available to aid recovery, different types of gambling products and the consequence of gambling and finally, money and sources of finance, as well as a miscellaneous topic, on the basis of the FREX words.

### **Negative emotions caused by gambling**

Topic one is the most common topic, appearing on average in around 25% of documents. Topic one relates to the emotional harm that gambling causes with a particular focus on the negative feelings associated with gambling including words like hurt, ashamed, deserve, hate, scared and ruin. This interpretation is supported by viewing the forum threads that are most representative of this topic. Note that to protect members' identity we paraphrased and anonymised the messages. The most representative post features a member talking about their feelings of mental anguish caused by addiction, the feeling of pain and a sense of hopelessness regarding how they can overcome their gambling problems.

[Insert Table 1]

### **Resources available to aid recovery**

Resources available to aid recovery can be further divided into formal and informal forms of



help with topics two and three reflecting more formal forms of help whereas topics four and five seem to illustrate self-help and peer-to-peer support for gambling problems. In total these four topics appear in ~40% of documents (topic two: 6%, topic three: 6%, topic four 4%, and topic five: 24%)

Topic two seems to relate to the use of Gordon Moody's inpatient services with FREX terms including Gordon, rehab, Moodies, Dudley, Beckenham, and programme. Gordon Moody has residential units in both Dudley and Beckenham. The most representative post from topic two mentions their use of the facilities at Gordon House during their recovery from compulsive gambling. High FREX words for topic three were therapy, GamCare, support, recovery, group, residential, recovering, problem etc. GamCare is a UK-wide helpline for people concerned about their gambling. The most representative post comes from the Gambling Therapy team informing users of the site about new groups and support that will become available to them and the second most representative post was of a gambler looking to find support from fellow forum users for a campaign they have started aiming to maintain provision of gambling support services being threatened with closure.

Topics four and five reflect more informal self-help and peer-to-peer support resources to help with recovery. Topic four contains a number of terms that relate to the process of recovery with terms like resilience, heal, advice, outlet, accept, and change. The top three most representative posts all revolve around users of the site posting advice they have come across, for example, on letting go and on dealing with the past. Topic five contains terms that relate to the use of the online forum as a way of seeking help and gaining hope, for example, FREX words are: post, forum, thread, hope, help, and thanks. The most representative post was of a member encouraging other members to post positive commitments that they want to make in their lives and the positive actions they will/have taken to make those commitments a reality.

[Insert Figure 3]

### **Different types of gambling products and the consequences of gambling**

Topics six and seven, which appear in approximately 15% and 8% of threads, respectively, contained posts where users talked about the different forms of gambling they had engaged in. Topic

six contains terms that relate to traditional high street gambling in the UK, especially bookmakers with high FREX words like bookies, football, roulette, bookmakers, fruit, and fobt (Fixed Odds Betting Terminal, a form of electronic gaming machine). Topic seven seems to relate to forms of betting and gaming with a more international focus, such as poker, blackjack, casino, trading, and stocks. The posts most representative of Topics six and seven all relate to gamblers who are giving accounts of how they gamble/d and the effect it has had on their life. For example, the most representative post of Topic six features a member who started playing fruit machines in early adulthood before moving on to arcades, casinos and online gambling with the results being considerable debts. The first most representative post from topic seven chronicles how a member started gambling before the age of 18 using free online sport betting products which gave them a false sense of security that losing bets is not problematic. As soon as they became legally able to gamble they joined an online website and began a roller coaster journey that ended up in debt. The second most representative post from Topic seven chronicles a member who started gambling in casino's before moving to slot machines in online casino's and eventually ending up losing considerable amounts of money playing blackjack.

### **Money and sources of finance**

Topic eight, which appears on average in 11% of the threads, contains a number of terms that suggest that users are discussing the issue of money with a particular focus on debt and ways of acquiring money so, for example, FREX words are: money, pay, credit, loan, bill, salary, payment, and debt. Interestingly, there are a number of family-related words like mom, sister, parent, and friend which suggests close relatives and friends have at times been a source of financial support. This idea is supported in the most representative post that features a member who had sold personal possessions to raise money for gambling which family and friends subsequently re-purchased for the member only for the items to be sold again to fund further gambling.

### **Miscellaneous**

There were additionally two topics (nine and ten) that appeared in a small percentage of documents (topic nine: ~4%; topic ten: ~2%) and do not appear to have themes that are clearly

meaningful to gambling. Instead, these consisted of commonly used words for example, didn't, don't, race, resent, spiritual etc. Topic models often contain themes that consist of common, miscellaneous words (AlSumait, Barbará, Gentle, & Domeniconi, 2009). These difficult to interpret topic can also arise in the trade-off between a model that has a better overall fit with higher held-out likelihood and lower residuals yet contain one or two topics, like topics nine and ten, that are more difficult to interpret (Chang, Gerrish, Wang, & Blei, 2009).

### **Discussion**

The topic modelling identified ten topics which can be best understood in terms of four overarching themes: negative emotions caused by gambling, resources to aid recovery, gambling products and consequences, and money and financing gambling, as well as a couple of miscellaneous themes.

The most commonly occurring topic involved discussion of the negative emotions caused by gambling. An interesting finding within these negative emotions was the sense of pain or hurting experienced by those using the forum. Some users wanted to numb their emotions, while others wanted to alleviate their feelings of pain. Previous work by Andrea, (2015) and Mudry and Strong (2013) have not identified such primary responses to gambling but have focussed on more self-conscious reflective emotions like shame and guilt. While some gamblers might be vulnerable to these feelings prior to gambling (Blazczynski & Nower, 2002), and certainly these feelings are common as they were observed among these topics, this study goes further as it allows us to infer these more primary emotion of pain and hurting are caused by the lived experience of disordered gambling.

The second theme concerned the resources available to aid recovery, comprising formal (e.g. residential facilities, support groups) and informal resources like the use of online forums. The most representative posts from topics four and five highlighted how the online forum is being used in imaginative ways by members to help, support and aid the recovery of others. In some cases, it was through sharing online content. In others, it was by encouraging others to make positive changes in their behaviour. These topics identify the features peculiar to online forums that may facilitate

recovery (i.e. asking members to post about positive commitments they are willing to make or sharing content they believe will be helpful to others). These findings parallel some of the change strategies identified by Rodda et al., (2018) who noted how some online gambling forum users seek inspiration from success stories posted by fellow forum users.

The third common theme was of gambling products, with topic six representing traditional venue-based forms of gambling whilst topic seven contains forms of gambling increasingly likely to be played online like playing poker, blackjack and betting on sports events. Both of these topics seem to reflect posts where members were describing their gambling journey in terms of the betting products they began with and how they moved onto other, in some cases, riskier, products. It is of particular interest that this is focused upon UK high street bookmakers, as these have been under intense scrutiny in relation to FOBTs, a product specifically mentioned in this topic. The types of products and the progress from one gambling product to another has not been reported in other research looking at gamblers online forum use. Identifying these products might be helpful to regulatory bodies, particularly if some of the products discussed (i.e. online gambling) are put forward for liberalisation.

Another novel finding from the correlated topic model is topic eight which seems to reflect that a number of posts talk about money and sources of finance they have used to fund their gambling problems. These posts have then typically also mentioned family and friends who have either provided financial support or suffered as a consequence of the gambling activities. Research from the other perspective of the family member and friends affected by a significant others gambling use support the idea that it can have substantial impact on their emotions, relationships and finances (Rodda, Lubman, & Dowling, 2017b). Indeed there is evidence to suggest that the harms caused by one person can impact on the lives of between 4-6 other people be it partners, parents, offspring or friends (Goodwin, Browne, Rockloff, & Rose, 2017). The eighth topic is also unusual insofar as it is the only topic that maps solely onto one of the DSM criteria for Gambling Disorder (American Psychiatric Association, 2013), namely relying on others (i.e. family, friends) for money to alleviate financial problems caused by gambling. Otherwise, we find some overlap between the criteria used in

measurements of disordered gambling (e.g. American Psychiatric Association, 2013; Ferris & Wynne, 2001; Lesieur & Blume, 1987; American Psychiatric Association, 2000) and the topics raised by gamblers who are experiencing addiction-related problems that are implicit in the text. Forum users report feelings of distress or irritability in regard to their gambling, difficulties in cutting back on their gambling, and preoccupation with gambling. What is interesting however is the absence of other life events that may emerge from gambling, such as committing illegal acts (the topic on financing tends to focus on licit activities such as obtaining credit), lying (which may be associated with feelings of shame and guilt, although speculating on this is beyond the scope of this paper), or jeopardising major opportunities.

These findings have the potential to inform psychoeducational programs at different stages of a gambling addiction. Using a behavioural insights framework (e.g. Michie, van Stralen, & West, 2011), educational interventions most effectively target psychological capability (i.e. knowledge, skills such as resilience), and reflective motivation (i.e. goals and plans). By distilling the content from the forums, it is possible to identify the themes that target these processes, in addition to identifying the language that gamblers are likely to perceive as relevant and timely. Subject to further research integrating them with existing programmes, content from the themes around resources to aid recovery, negative emotionality and gambling related products (and consequences) are good candidates for this purpose. These findings identify a few areas and indeed specific types of resources where a programme might be beneficial in order to build capacity among gamblers. Previous research has shown that these forums have a much wider audience than the people who post on them (Cooper, 2004). Because very few people who meet the criteria for gambling disorder engage with programs designed to help them (Melville et al., 2007), there remains a critical untapped need for resources for people who need help but do not feel willing to engage with others about their addiction.

Following on from this, these findings could form the basis of a brief digital intervention to help gamblers at moments of crisis, or hazardous use (e.g. Nower & Blaszczynski, 2003), especially as most gamblers will seek self-directed care as a first instance. In other addictive behaviours, such as hazardous drinking, it has been shown that brief interventions based on monitoring engagement with

an addictive behaviour, such as logging alcohol consumption, as well as reflective motivations such as goal setting or planning, are effective at mitigating hazardous behaviour (Kaner et al., 2017). The findings of this study might be utilised for a similar purpose, especially as there is promising evidence that the same processes that would be maximally effective in a psychoeducational intervention would be similarly efficacious here. Such an intervention could include logging gambling behaviour (e.g. type activity, length of play, amount spent), alongside the life events (e.g. consequences of gambling on oneself and their family, having to source extra forms of credit) and emotional impacts that might emerge from gambling. It could also include information that could provide sources of formal and informal support as well drawn from the themes that emerged here, so that users have sufficient capacity to address their problems if needed. The rich source of textual data also allows the possibility of segmenting users and tailoring content to match user needs, by cross-comparing it with data from webscraped sources such as these online forums.

### **Limitations.**

There are a number of potential limitations to be aware of with this analysis. It is not clear how representative forum users are, especially of the wider population of people with gambling-related problems. It has been frequently noted that a minority of people with addiction related problems seek treatment or participate in addiction research (Melville, Casey, & Kavanagh, 2007; Susukida, Crum, Stuart, Ebnesajjad, & Mojtabai, 2016). Because two of the themes relate to the use of residential programmes, it appears that some gamblers using the forum have engaged with treatment services. Therefore, it is unclear in this specific instance whether scraping this forum is capturing a distinct group of gamblers from the existing literature. In addition, it might be the case that such data might be confounded by the most frequent users also being likely to have the greatest problems. However, we attempted to control for this by focusing on the initial post within each thread. The data are from initial forum posts, and so it is unclear whether posts that discussed certain topics were associated with more successful attempts to control gambling. Although we scraped all of the posts from the gambling forum, some content might have been removed by moderators for violating

the rules of the forum. It cannot be ascertained what section of the gambling community is being captured by the web scraping, although it is likely to be predominantly (but not limited to) people experiencing problems in the UK. Although this captures a breadth of gambling activities, further topics may be identified by scraping forums devoted to specific gambling activities.

### **Conclusions.**

Correlated topic modelling of web scraped forum posts revealed ten topics that could be organised into four themes: dealing with negative emotions caused by gambling, recovery and treatment, gambling products and borrowing money. These findings identify the topics that gamblers spontaneously raise when talking about their problems, and the language used to express these. The content of the topics, subject to successful evaluation, might be integrated with existing interventions, either in terms of the language used to make it more relevant and timelier to gamblers, or the content. Further research ought to evaluate the robustness of these themes across other samples of gamblers (e.g. samples with known gambling disorder, other gambling forums, Twitter), and to explore them in more depth among groups of engaged gamblers.

## References

- Adjerid, I., & Kelley, K. (2018). Big Data in Psychology: A Framework for Research Advancement. *American Psychologist*. <https://doi.org/10.1037/amp0000190>
- Airoldi, E. M., & Bischof, J. M. (2016). Improving and Evaluating Topic Models and Other Models of Text. *Journal of the American Statistical Association*, *111*(516), 1381–1403. <https://doi.org/10.1080/01621459.2015.1051182>
- AlSumait, L., Barbará, D., Gentle, J., & Domeniconi, C. (2009). Topic significance ranking of LDA generative models. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *5781 LNAI(PART 1)*, 67–82. [https://doi.org/10.1007/978-3-642-04180-8\\_22](https://doi.org/10.1007/978-3-642-04180-8_22)
- American Psychiatric Association. (2000). *Diagnostic Criteria from the DSM-IV*. Washington, DC: American Psychiatric Association.
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders*. Washington, DC: Author. <https://doi.org/doi.org/10.1176/appi.books.9780890425596>
- Andrea, C. (2015). Sharing problem gamblers' experiences: A text analysis of gambling stories via online forum. *Mediterranean Journal of Clinical Psychology*, *3*(1), 1–26. <https://doi.org/10.6092/2282-1619/2015.3.1051>
- Blei, D. M., Edu, B. B., Ng, A. Y., Edu, A. S., Jordan, M. I., & Edu, J. B. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, *3*, 993–1022. <https://doi.org/10.1162/jmlr.2003.3.4-5.993>
- Blei, D. M., & Lafferty, J. D. (2007). A Correlated Topic Model of Science. *Source: The Annals of Applied Statistics*, *1*(1), 17–35. Retrieved from <http://www.jstor.org/stable/4537420>
- Brown, J., West, R., Angus, C., Beard, E., Brennan, A., Drummond, C., ... Michie, S. (2016). Comparison of brief interventions in primary care on smoking and excessive alcohol consumption: A population survey in England. *British Journal of General Practice*, *66*(642),



e1–e9. <https://doi.org/10.3399/bjgp16X683149>

Chang, J., Gerrish, S., Wang, C., & Blei, D. M. (2009). Reading Tea Leaves: How Humans Interpret Topic Models. *Advances in Neural Information Processing Systems* 22, 288–296.

<https://doi.org/10.1.1.100.1089>

Cooper, G. (2004). Exploring and Understanding Online Assistance for Problem Gamblers: The Pathways Disclosure Model, *I*(2), 32–38.

Cowlshaw, S., Gale, L., Gregory, A., McCambridge, J., & Kessler, D. (2017). Gambling problems among patients in primary care: A cross-sectional study of general practices. *British Journal of General Practice*, 67(657), e274–e279. <https://doi.org/10.3399/bjgp17X689905>

Ferris, J., & Wynne, H. (2001). The Canadian Problem Gambling Index : Final report. *Canadian Centre on Substance Abuse*, 38. <https://doi.org/10.1007/s10899-010-9224-y>

Gainsbury, S., & Blaszczynski, A. (2011a). Online self-guided interventions for the treatment of problem gambling. *International Gambling Studies*, 11(3), 289–308.

<https://doi.org/10.1080/14459795.2011.617764>

Gainsbury, S., & Blaszczynski, A. (2011b). Online self-guided interventions for the treatment of problem gambling. *International Gambling Studies*, 11(3), 289–308.

<https://doi.org/10.1080/14459795.2011.617764>

Goodwin, B. C., Browne, M., Rockloff, M., & Rose, J. (2017). A typical problem gambler affects six others. *International Gambling Studies*, 17(2), 276–289.

<https://doi.org/10.1080/14459795.2017.1331252>

Grimmer, J., & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, 21(3), 267–297.

<https://doi.org/10.1093/pan/mps028>

Grün, B., & Hornik, K. (2011). topicmodels: An R Package for Fitting Topic Models. *Journal of Statistical Software*, 40(13), 1–30. Retrieved from <https://cran.r->

project.org/web/packages/topicmodels/vignettes/topicmodels.pdf

Kaner, E., Beyer, F., Garnett, C., Crane, D., Brown, J., Muirhead, C., ... Michie, S. (2017).

Personalised digital intervention for reducing hazardous and harmful alcohol consumption in community dwelling populations (Review). *Cochrane Database of Systematic Reviews*, (9), 1–159. <https://doi.org/10.1002/14651858.CD011479.pub2>. [www.cochranelibrary.com](http://www.cochranelibrary.com)

Langham, E., Thorne, H., Browne, M., Donaldson, P., Rose, J., & Rockloff, M. (2016).

Understanding gambling related harm: A proposed definition, conceptual framework, and taxonomy of harms. *BMC Public Health*, 16(1). <https://doi.org/10.1186/s12889-016-2747-0>

Lesieur, H., & Blume, S. (1987). The South Oaks Gambling Screen (SOGS): A new instrument for the identification of pathological gamblers. *The American Journal of Psychiatry*, 144(9), 1184–1188.

Melville, K. M., Casey, L. M., & Kavanagh, D. J. (2007). Psychological treatment dropout among pathological gamblers. *Clinical Psychology Review*, 27(8), 944–958. <https://doi.org/10.1016/j.cpr.2007.02.004>

Michie, S., van Stralen, M. M., & West, R. (2011). The behaviour change wheel: a new method for characterising and designing behaviour change interventions. *Implementation Science : IS*, 6(1), 42. <https://doi.org/10.1186/1748-5908-6-42>

Mimno, D., Wallach, H. M., Talley, E., Leenders, M., & Mccallum, A. (2011). Optimizing Semantic Coherence in Topic Models, 262–272. Retrieved from [http://delivery.acm.org/10.1145/2150000/2145462/p262-mimno.pdf?ip=128.243.2.30&id=2145462&acc=OPEN&key=BF07A2EE685417C5.9530128D756F5CF.4D4702B0C3E38B35.6D218144511F3437&CFID=1006212197&CFTOKEN=36947674&\\_\\_acm\\_\\_=1510756881\\_cdacd39081cd51c8a5c43d7d22f38b76](http://delivery.acm.org/10.1145/2150000/2145462/p262-mimno.pdf?ip=128.243.2.30&id=2145462&acc=OPEN&key=BF07A2EE685417C5.9530128D756F5CF.4D4702B0C3E38B35.6D218144511F3437&CFID=1006212197&CFTOKEN=36947674&__acm__=1510756881_cdacd39081cd51c8a5c43d7d22f38b76)

Mudry, T. E., & Strong, T. (2013). Doing recovery online. *Qualitative Health Research*, 23(3), 313–325. <https://doi.org/10.1177/1049732312468296>

- Nower, L., & Blaszczynski, A. (2003). Binge Gambling: A Neglected Concept. *International Gambling Studies*, 3(1), 23–35. <https://doi.org/10.1080/14459790304589>
- Paxton, A., & Griffiths, T. L. (2017). Finding the traces of behavioral and cognitive processes in big data and naturally occurring datasets. *Behavior Research Methods*, 49(5), 1630–1638. <https://doi.org/10.3758/s13428-017-0874-x>
- Roberts, M. E., Stewart, B. M., & Tingley, D. (n.d.). stm: R package for Structural Topic Models. *Journal Of Statistical Software*. <https://doi.org/10.18637/jss.v000.i00>
- Roberts, M. E., Stewart, B. M., & Tingley, D. (2015). Navigating the Local Modes of Big Data: The Case of Topic Models \*. Retrieved from <https://scholar.harvard.edu/files/dtingley/files/multimod.pdf>
- Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., ... Rand, D. G. (2014). Structural Topic Models for Open-Ended Survey Responses. *American Journal of Political Science*, 58(4), 1064–1082. <https://doi.org/10.1111/ajps.12103>
- Rodda, S., Lubman, D., & Dowling, N. (2017a). *Examining the impact of e-mental health in problem gambling*. Retrieved from [https://www.responsiblegambling.vic.gov.au/\\_\\_data/assets/pdf\\_file/0003/33519/Rodda\\_examining\\_impact\\_e-mental\\_health\\_problem\\_gambling\\_Jan\\_2017.pdf](https://www.responsiblegambling.vic.gov.au/__data/assets/pdf_file/0003/33519/Rodda_examining_impact_e-mental_health_problem_gambling_Jan_2017.pdf)
- Rodda, S., Lubman, D., & Dowling, N. (2017b). *Online support needs and experiences of family members affected by problem gambling*. Retrieved from <https://aifs.gov.au/agrc/sites/default/files/publication-documents/agrc-rr-online-support-web.pdf>
- Rodda, S. N., Hing, N., Hodgins, D. C., Cheetham, A., Dickins, M., & Lubman, D. I. (2018). Behaviour change strategies for problem gambling: an analysis of online posts. *International Gambling Studies*. <https://doi.org/10.1080/14459795.2018.1432670>
- RStudio Team. (2015). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA: RStudio, Inc. Retrieved from <http://www.rstudio.com/>

Susukida, R., Crum, R. M., Stuart, E. A., Ebnesajjad, C., & Mojtabai, R. (2016). Assessing sample representativeness in randomized controlled trials: application to the National Institute of Drug Abuse Clinical Trials Network. *Addiction, 111*(7), 1226–1234.

<https://doi.org/10.1111/add.13327>

Wickham, H. (2015). rvest: Easily harvest (scrape) web pages.

Wood, R. T. A., & Wood, S. A. (2009). An evaluation of two United Kingdom online support forums designed to help people with gambling issues. *Journal of Gambling Issues, 0*(23), 5–30.

Retrieved from <http://jgi.camh.net/index.php/jgi/article/view/3807/3818>

## Tables

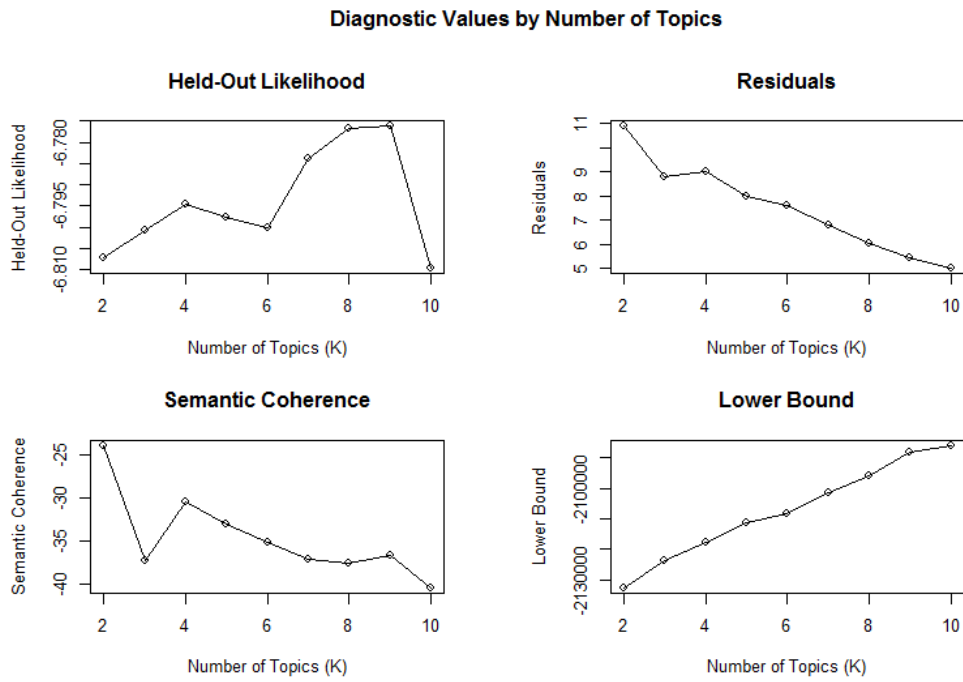
*Table 1.*

Top twenty highest Frex words per topic.

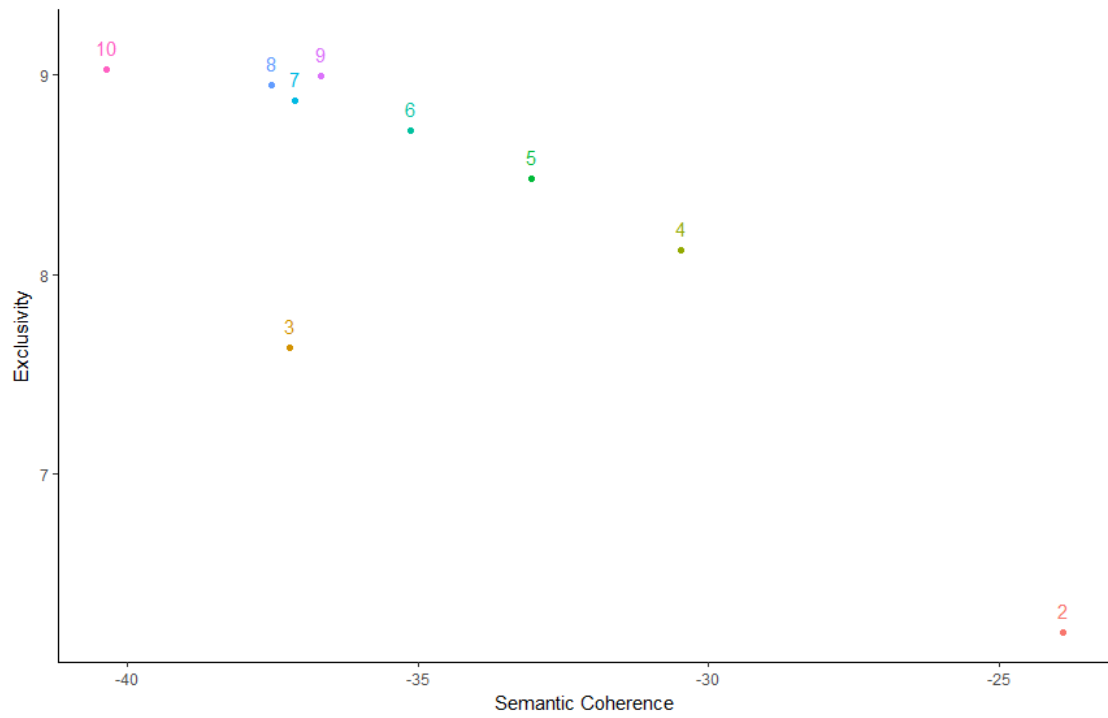
Theme	Topic Number	Top twenty highest Frex words (including stem words)
<b>Negative emotions caused by gambling</b>	Topic 1	dont, husband, cant, doesnt, want, hurt, feel, anymor, sick, that, tonight, asham, know, deserv, ruin, hate, mayb, scare, tire, right
	Topic 2	gordon, rehab, prison, your, daughter, moodi, ticket, beckenham, slip, wasnt, whilst, realis, budget, nervous, week, hous, dudley, job, programm, met
<b>Resources available to aid recovery</b>	Topic 3	group, therapi, gamcar, member, address, 've, maverick, residenti, helpin, servic, 're, moodi, support, pressur, associ, recoveri, reflect, gordon, trigger, honesti
	Topic 4	resili, heal, rock, path, past, arent, advis, need, outlet, aunti, attach, label, stage, issu, pictur, allow,*, accept, memori, give
	Topic 5	post, free, relaps, everyon, forum, hope, thread, read, christma, hello, compuls, thank, site, today, journal, journey, quit, day, join, new
<b>Different types of gambling products and consequences</b>	Topic 6	booki, ive, footbal, roulett, machin, wage, spin, pound, mum, bookmak, partner, win, overdraft, quid, shop, xmas, bet, fruit, fobt, exclud
	Topic 7	poker, sport, deposit, blackjack, game, play, dollar, onlin, player, bet, profit, trade, loss, stock, won, bonus, small, chase, amount, wager
<b>Money and sources of finance</b>	Topic 8	parent, pay, credit, casino, loan, sister, bill, euro, mom, borrow, school, salari, card, went, payment, told, debt, sold, paid, money
<b>Miscellaneous</b>	Topic 9	spiritu, resent, season, *, forgiv, edit, *, law, activ, holiday, grate, exercis, water, practic, sober, *, independ, celebr, blue, special
	Topic 10	didn't, race, don't, hors, dream, hospit, wasn't, doctor, couldn't, missus, won't, diseas, inpati, don, phase, funni, episod, ugly, steal, can't

Note 1. \* represents common forenames removed to protect forum users identity.

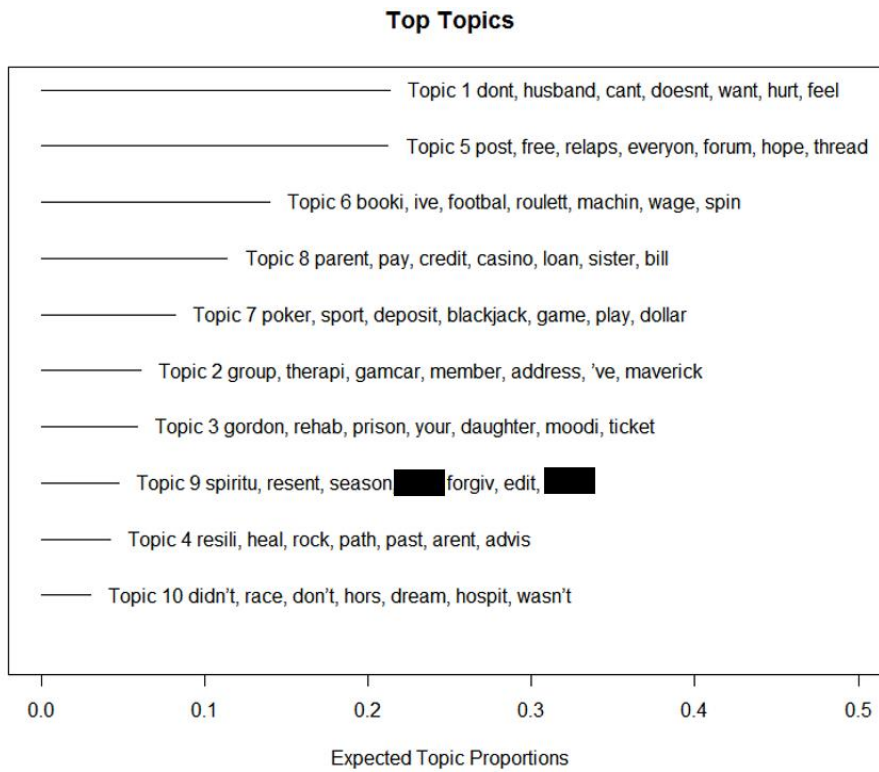
## Figures



*Figure 1.* Diagnostic values by the number of topics. Greater held-out likelihood, smaller residuals, smaller semantic coherence and smaller lower bound are indicative of better fitting models for each metric.



*Figure 2.* Each model plotted on their average exclusivity and semantic coherence scores across topics. Semantic coherence is a metric of topic quality. Exclusivity refers to the extent to which the most frequent words in a topic are unique to that topic.



*Figure 3.* Show the top seven Frex words per topic and the average topic proportions across all the documents included in the model (topics ordered by expected proportions). Two words are blanked out for to protect anonymity (common forenames).