

**Integrating computerized linguistic and social network analyses to capture addiction recovery capital in an online community**

BLIUC, Ana-Maria, IQBAL, Muhammad and BEST, David  
<<http://orcid.org/0000-0002-6792-916X>>

Available from Sheffield Hallam University Research Archive (SHURA) at:

<http://shura.shu.ac.uk/24903/>

---

This document is the author deposited version. You are advised to consult the publisher's version if you wish to cite from it.

**Published version**

BLIUC, Ana-Maria, IQBAL, Muhammad and BEST, David (2019). Integrating computerized linguistic and social network analyses to capture addiction recovery capital in an online community. *Journal of visualized experiments : JoVE* (147).

---

**Copyright and re-use policy**

See <http://shura.shu.ac.uk/information.html>

## Video Article

# Integrating Computerized Linguistic and Social Network Analyses to Capture Addiction Recovery Capital in an Online Community

Ana-Maria Bliuc<sup>1</sup>, Muhammad Iqbal<sup>2</sup>, David Best<sup>3</sup><sup>1</sup>School of Social Sciences and Psychology, Western Sydney University<sup>2</sup>Institute for Sustainable Industries and Liveable Cities (ISILC), Victoria University<sup>3</sup>Centre for Regional Economic and Social Research, Sheffield Hallam UniversityCorrespondence to: Ana-Maria Bliuc at [A.Bliuc@westernsydney.edu.au](mailto:A.Bliuc@westernsydney.edu.au)URL: <https://www.jove.com/video/58851>DOI: [doi:10.3791/58851](https://doi.org/10.3791/58851)

Keywords: Behavior, Issue 147, online social interaction, online community, addiction, recovery, supportive networks, social identity

Date Published: 5/31/2019

Citation: Bliuc, A.M., Iqbal, M., Best, D. Integrating Computerized Linguistic and Social Network Analyses to Capture Addiction Recovery Capital in an Online Community. *J. Vis. Exp.* (147), e58851, doi:10.3791/58851 (2019).

## Abstract

The article describes a new methodology designed with the aim of finding a comprehensive, unobtrusive, and accurate way of capturing social recovery capital development in online communities of recovery from alcohol and drug (AOD) addiction. Recovery capital was conceptualised as both engagement in the online recovery community and identification with the community. To measure recovery capital development, naturally occurring data were extracted from the social media page of a specific recovery program, with the page being set up as a resource for a face-to-face recovery program. To map engagement with the online community, social network analysis (SNA) capturing online social interaction was performed. Social interaction was measured through the linkages between the online contributors/members of the online community as represented by program clients, staff, and supporters from the broader community. To capture markers of social identification with the online community, computerised linguistic analysis of the textual data (content from posts and comments) was conducted. Recovery capital captured in this way was analysed against retention data (a proxy outcome indicator), as days spent in the (face-to-face) recovery program. The online data extracted was linked to participant data in regards to program retention to test prediction of a key recovery outcome. This approach allowed the examination of the role of online support communities and assessment of the association between recovery capital (developed via the online community of recovery) and recovery outcomes.

## Video Link

The video component of this article can be found at <https://www.jove.com/video/58851/>

## Introduction

The presented method has been designed to capture alcohol and other drugs (AOD) addiction recovery capital in online contexts. In the field of addiction, recovery capital has been defined as "the sum total of one's resources that can be brought to bear on the initiation and maintenance of substance misuse cessation"<sup>1</sup>. Recovery capital has been primarily measured through self-reports<sup>2,3</sup> in face-to-face contexts. This approach provides an alternative method of measuring recovery capital in online contexts by capturing the quality and quantity of online interactions in online communities of recovery.

Given the steady increase in the use of online resources in the form of peer-support in a range of health-related issues<sup>4,5</sup>, it is necessary to develop new methods to capture the quality of these resources. Online peer support occurs in the form of social interactions in online forums and communities. Supportive social interactions in these online contexts contribute to building recovery capital, which in turn has a positive impact on the recovery process<sup>6,7</sup>. The method proposed presents a number of advantages over alternative methods. Firstly, it overcomes some of the limitations involving the use of self-report measures in addiction research, particularly around recall and self-presentational biases. While self-report measures are considered to have reasonable levels of reliability and validity, they are susceptible to biases and inaccuracies. To enhance accuracy and minimize bias, it has been recognized that there is a need to increase the use of novel measures and data collection situations designed to avoid or minimize these issues<sup>8</sup>. By accessing data naturally occurring in contexts where people in various stages of recovery interact spontaneously, and by using analysis methods that can extract meaningful information from these data (able to capture indicators of psychological states), biases due to social desirability (self-presentational) and inaccuracies due to limitations in recall can be reduced or even eliminated. Secondly, this method is highly efficient and cost-effective, as it relies on the extraction of already existing online data (i.e., in open online forums that are publicly accessible).

Described next is the method that was applied to a study of building recovery capital in an online community established to complement a traditional, face-to-face addiction recovery program for addicts in early recovery stages. In this case, online (social media) data were linked to program retention data, but the method can be also used in cases where linkage data is not available or accessible.

## Protocol

The research described here was approved by the research ethics community at Sheffield Hallam University.

### 1. Setup

NOTE: Please refer to the attached R script provided as **Supplementary File 1**.

1. Load required packages (Rfacebook<sup>9</sup>, dplyr<sup>10</sup>, igraph<sup>11</sup>, and openxlsx<sup>12</sup>) in R. Packages refer to functions, datasets, or compiled code that allow users to analyze, transform, or extract data.
2. Load (external) retention and user data into R as a data frame from a CSV file.  
NOTE: Retention data refers to the number of days in which a client participates in the offline (traditional) addiction recovery program. It was provided by the administrator of the (offline) recovery program as recorded onto a CSV file with the participant name and number of days they have been involved in the program. The participant name was replaced by the anonymous ID number prior to being imported into R.

### 2. Data extraction from the online community (the social page of an addiction recovery community)

NOTE: This protocol applies to a social media page, but it can be adapted to different types of online communities. In the case of the Rfacebook package, it allows the user to extract data from the social media page into R.

1. Create a social media (Facebook) access token by following the guide on the referenced website<sup>13</sup>.
2. Create access token in R.
3. Using the "getGroup" function from Rfacebook, extract data from the social media page of the community of interest (e.g., content of post, number of comments and likes for each post, a unique ID number for each post, etc.). This data is then saved as a data frame.  
NOTE: A data frame is essentially a table within R used to store data.
4. Using the "getPost" function from Rfacebook, along with the Post IDs extracted in step 2.3, extract data about posts likes made on the page.
5. Using the "getPost" function from Rfacebook, along with the Post IDs extracted in step 2.3, extract data on the comments made on each post (e.g., user IDs of people commenting the post, when the comment was made, how many likes the post received). This data is then saved as a data frame.
6. Using the comment IDs extracted in step 2.5, extract data on the "comment likes" made on each post (e.g., user IDs of people liking the comment). This data is then saved as a data frame.
7. Combine the posts, post likes, comments, and comment likes data into one data frame.
8. Add a monthly breakdown (i.e., month 1 to 8).

### 3. Calculation of social media activity made and received by each client

1. Calculate the number of posts, comments, post likes and comment likes made by each client.
2. Calculate the number of posts, comments, post likes and comment likes received by each client.
3. Join the data frame of social media activity made and received by each client to the retention data frame.
4. Calculate the difference between posts and comments with likes and no likes.
5. Calculate the difference between posts with comments and no comments.
6. Join the likes difference data to the retention data.
7. Join the comments difference data to the retention data.
8. Calculate all the likes made by each client.
9. Calculate all the likes received by each client.
10. Identify which users did not participate in social media group (i.e., no activity).

### 4. Conducting social network analysis

1. Create an edge list. An edge list is a list of relationships within the social network, which in this case is based on 1) liking posts and comments and 2) commenting on posts. This is done by looking at two columns within the dataset. The first column contains the anonymous ID of the person making the post, while the second contains the anonymous ID of the person liking or commenting on the post.
2. Create a vertex list. A vertex list is a list of all individuals in the group. This is done by converting the two columns in the list of relationships into one column, and removing duplicate anonymous IDs so only the unique anonymous ID is left.
3. Using the "graph.data.frame" and "get.adjacency" functions in the igraph package, create graph and graph matrix objects from the edge and vertex lists.
4. Using the "degree" and "betweenness" functions from the igraph package, obtain the network statistics (degree and betweenness) of the online group.

### 5. Conducting computerized linguistic analysis in LIWC

1. Export textual social media data (i.e., posts and comments) and post/comment ID column into CSV files.
2. Import the CSV files of textual social media data into the Linguistic Inquiry Word Count (LIWC) software.

3. Generate LIWC categories and save to new CSV files. Do this by clicking on “Analyze Text”, then on “Excel/CSV file”, and clicking on the column containing the posts and comments to select the text to be analyzed. After LIWC has completed analyzing the textual data, save the output as a new CSV file.
4. Import the LIWC results CSV file into R, and merge with existing data. The data is matched by the post/comment ID column, which exists in both LIWC and existing data frames.
5. Calculate total LIWC scores for each user in posts and comments, then join to the retention data.
6. Calculate total LIWC scores for each user in all textual data (post and comments combined), then join to the retention data.
7. Remove NAs from the retention data data frame.

## 6. Conducting regression analysis (to determine if indicators of engagement with the online community predict retention in the offline recovery program)

1. Define the independent variables.
2. Using the “lm” function in base R, conduct linear regression analysis using the retention data as the dependent variable, and LIWC categories, comments, post likes, and comment likes as independent variables.
3. Combine regression analysis results into one data frame.

## 7. Creating monthly SNA maps

1. Prepare data frames for SNA Maps.
2. Create an edge list based on monthly cumulative social media activity.
3. Create a vertex list based on monthly cumulative social media activity.
4. Create graphs and graph matrices based on monthly cumulative social media activity.
5. Set the layout of SNA maps based on cumulative social media activity.
6. Add colors based on user roles.
7. Create SNA maps and save them to a file.

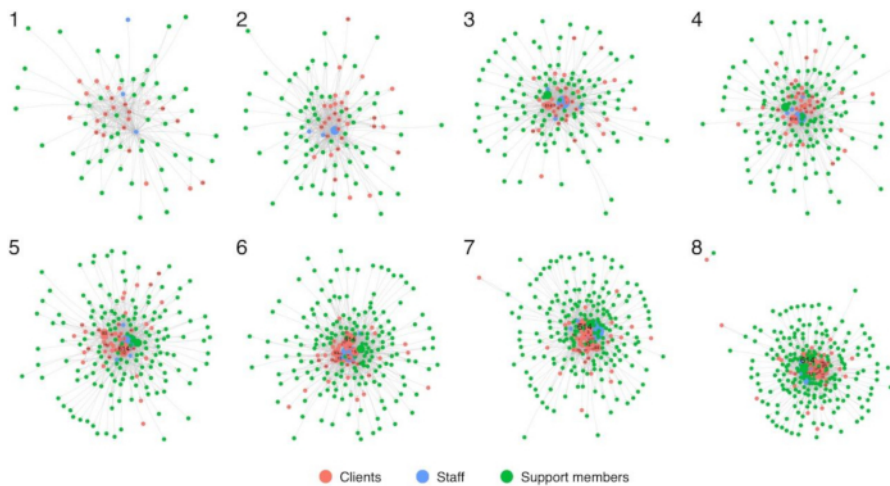
## 8. Calculating monthly cumulative social media activity of the social media group

1. Calculate monthly cumulative social media activity by staff, clients, and other members of the social media group.
2. Calculate monthly cumulative social media activity by all members of the social media group.
3. Join the monthly cumulative social media activity data frames together.

## Representative Results

A detailed description of representative results obtained using this method can be found in our recent work<sup>14</sup>, which was reviewed and received full approval from the research ethics committee of the institution at which the research was conducted. In the report described here, the study investigated whether online participation in a community of recovery contributes to the recovery process through recovery capital building (as captured by increased levels and quality of online social interactions and positive identity development). In other words, the study examined whether indicators of online recovery capital developed over the eight months of online data assessed and also predicted retention in a recovery program designed for fostering community involvement for addicts in early stages of recovery.

To map how participants interacted online, social network analysis (SNA) using data extracted from the social media page ( $n = 609$ ) of a recovery community was conducted. A visual representation of the social network and its evolution is presented in **Figure 1**. The figure illustrates the activity in the online community observed each month for a period of 8 months in the form of connections between all participants in the online community (i.e., commenting on posts, liking posts, and liking comments). The number of connections that an “agent” in the network has determines how central they will be in the social network. Computerized linguistic analysis was used to assess the textual data (capturing social identity markers), and linear regression analysis was conducted to determine whether the indicators of recovery capital predicted program retention. These analyses indicated that program retention was indeed predicted by: (a) levels of group validation received in the form of comment likes and all likes received on the social media page, (b) position in the social network (network centrality), and (c) group identity and achievement (as captured by the linguistic content of online communication). The results supported the argument that, overall, positive social interactions between members of an online recovery community are supportive of the recovery process. A summary of those findings is presented below.



**Figure 1: Monthly representations of the social network of the online community over 8 months suggest changes in the pattern of social interactions between the participants.** These representations illustrate how at the start, most of the client members in the online community (clients of the offline recovery program) are mostly disconnected, and it is the program staff and only a small number of clients who drive the online activity. However, this gradually changes, so that after 8 months, the clients are the ones most connected (therefore the most central), with the highest number of connections in the network (figure is adapted from a previous publication)<sup>14</sup>. [Please click here to view a larger version of this figure.](#)

#### Descriptive statistics

Participants' levels of engagement with the online community were measured by computing the contributions of all participants in the online community as number of posts, comments, and likes made by staff, clients, and broader community members. **Table 1** presents a breakdown by type of contribution (as made by each category of participant) across 8 months.

Group members	Type of online contribution	Month 1	Month 2	Month 3	Month 4	Month 5	Month 6	Month 7	Month 8
All	Posts and comments	382	388 (770)	579 (1349)	369 (1718)	530 (2248)	581 (2829)	796 (3625)	674 (4299)
	Post likes given	1167	878 (2045)	1856 (3901)	1440 (5341)	1880 (7221)	1756 (8977)	2667 (11644)	1857 (13501)
	Comment likes given	784	970 (1604)	825 (2429)	171 (2600)	634 (3234)	970 (4204)	825 (5029)	171 (5200)
Staff	Posts and comments	129	106 (235)	170 (405)	96 (501)	185 (686)	176 (862)	227 (1089)	316 (1405)
	Post likes given	188	147 (335)	302 (637)	209 (846)	385 (1231)	372 (1603)	567 (2170)	511 (2681)
	Comment likes given	168	303 (471)	237 (708)	69 (777)	168 (945)	303 (1248)	237 (1485)	69 (1554)
Clients	Posts and comments	145	155 (300)	214 (514)	132 (646)	208 (854)	286 (1140)	419 (1559)	253 (1812)
	Post likes given	365	252 (617)	415 (1032)	303 (1335)	549 (1884)	529 (2413)	898 (3311)	576 (3887)
	Comment likes given	143	318 (461)	235 (696)	33 (729)	143 (872)	318 (1190)	235 (1425)	33 (1458)
Others	Posts and comments	108	127 (235)	195 (430)	141 (571)	137 (708)	119 (827)	150 (977)	105 (1082)
	Post likes given	614	479 (1093)	1139 (2232)	928 (3160)	946 (4106)	855 (4961)	1202 (6163)	770 (6933)
	Comment likes given	473	349 (672)	353 (1025)	69 (1094)	323 (1417)	349 (1766)	353 (2119)	69 (2188)

**Table 1: Shown is the number of online contributions by type (post and comments made, likes given to posts, and likes given to comments) by members of the online community on the across 8 months.** The members of the online community are classified as staff (support staff employed by the offline recovery program), clients (people in recover who are participating in the offline recovery program), and others (supporters and pro-recovery advocates from the broader community).

**Determinants of retention in the program**

The following hypotheses were tested: (1) program retention should be associated with indicators of recovery capital development (i.e., reflected in the quantity and quality of online interaction), and (2) program retention should also be associated with indicators of identity change, (i.e., indicators of positive recovery identity development). The quantity of online interaction was indicated by the a) number of posts made, b) number of comments made, c) number of post likes received, d) number of comment likes received, and e) number of all likes received.

To determine the quality of online interaction, network structure and language content were analyzed. More specifically, degree and betweenness coefficients derived from social network analysis (SNA) and linguistic indicators of positive affect derived from computerized linguistic analysis were used. As indicators of positive identity change (as identification with the recovery community) the frequency of use of the pronoun “we” and achievement words (e.g., try, goal, win, etc.) were used. Finally, the dependent variable (retention in the program) was indicated by the total number of days spent in the program (ranging from 86 to 464 days here). As shown by the results, levels of online interaction and in-group validation (as reflected by the number of likes received for posts and comments) predicted program retention (**Table 2**). Program retention was also predicted by identification markers (as captured by the use of the pronoun “we” in posts and of achievement words in both posts and comments). Finally, where participants are situated within the social network (i.e., degree of centrality) also represents an important aspect of retention (**Table 2**).

Variable	B	SE	$\beta$	R <sup>2</sup>
Comment likes received	0.43	0.18	.47*	0.22
Likes received (all)	0.08	0.03	.43*	0.18
Comment-like difference	1.09	0.5	.43*	0.19
Network degree	0.01	0	.43*	0.18
LIWC We (Post)	3.89	1.76	.43*	0.19
LIWC Achievement (Post)	0.56	0.26	.43*	0.18
LIWC Achievement (All)	0.14	0.07	.42*	0.17

**Table 2: Retention time as predicted by online engagement, network statistics, and linguistic categories.**

## Discussion

The approach described here is based on a new method of measuring how online group processes can impact retention in an addiction recovery program. Applying this method to an online community of recovery from addiction, it was found that there were four key aspects predicted program retention: being highly involved in the online community, being central in the online social network, positive affect expressed in communication with other members of the online community, and receiving validation from others for contributions to the network<sup>14</sup>. The findings obtained by using this method support existing theoretical models of recovery. That is, two key models in the recovery literature, the Social Identity Model of Recovery<sup>15</sup> and the Social Identity Model of Cessation Maintenance<sup>16</sup>, both emphasize the importance of active participation in groups which are supportive of recovery. Both models suggest that increased identification and commitment to such groups contribute to lower future contact with using groups and consequent relapse.

As illustrated in our research, the method allowed us to map out trajectories of recovery or change of individual members of the online community<sup>14</sup>. Visualizations of the online social networks and their evolution over time can provide valuable information about the movement of members of the online community from the periphery to the center of the network and vice-versa (these movements in the network indicate changes in levels of engagement with the online community). In a 2017 study<sup>14</sup>, interviews with members of the online community who undertook the most significant changes in terms of movement from the periphery to center of the networks were conducted as a way of triangulating our findings based on SNA, computerized linguistic analysis, and regression against retention data. Future studies may focus instead on those members who became disengaged with the online community, on those who never become engaged, or on more direct measures of outcome such as substance use and reoffending. This methodology can further fine-tuned to be used in intervention programs, for example, for assessing the role of moderators in help forums.

There are currently no studies providing evidence on the benefits of the method described here when used by itself (the method described was used in conjunction with retention data and triangulated with qualitative data from interviews with key online community members<sup>14</sup>), but this approach can provide accurate and bias-free data that can complement self-reporting and other measures in studies of addiction recovery.

This method was applied to examine online social interactions in the context of a social media page established as a complementary form of support to a standard, face-to-face recovery program. However, with minor changes, the method can be used to investigate online social interactions in other types of online communities (online forums, discussion groups, chat rooms, commentary websites, etc.). One of the key advantages of this method is that it can be adapted and applied to contexts beyond communities of addiction recovery to any online community. For example, in our own political psychology research, we use a similar method (developed from the method described here) to capture the quality of online interactions and changes in these interactions between members of far-right online communities. In effect, the method can be applied to any online community in which data in the form of connections between members (as social network linkages) and linguistic content can be extracted.

However, in accessing and working with online data, researchers need to be aware of ethical issues, some which apply to self-reporting and other types of data in general and some which are only encountered in an online environment. In the research described here (which was approved by the research ethics community at Sheffield Hallam University), consent was obtained from the organization managing the recovery program, and strict measures were taken to ensure complete anonymity of participants in the open social media page (e.g., after online and retention data matching, all identifying information was removed from the files and also no potentially self-identifying quotes were used from the publicly accessible online communication).

Close communication with the organization also ensured that the participants in the program were aware of the study and research findings, and one of the researchers met regularly with the group to explain the study and its results. In other cases, however, where online communities are not associated with specific offline programs, it may be harder to determine who should be asked for consent regarding data extraction (applicable especially in unmoderated forums, where people in recovery seek online peer support). While the general principles of ethical research will apply, researchers need to adopt a case-by-case approach to ensure that the extraction and analysis of online data does not pose any significant risks to the participants (e.g., compromising privacy).

## Disclosures

The authors have nothing to disclose.

## Acknowledgments

We are grateful to the clients and staff of Jobs, Friends and Houses, UK, who supported and agreed to participate in our research.

## References

1. Cloud, W., Granfield, R. Conceptualizing recovery capital: Expansion of a theoretical construct. *Substance Use and Misuse*. **43**, 1971-1986 (2008).
2. Best, D. et al. Mapping the recovery stories of drinkers and drug users in Glasgow: Quality of life and its associations with measures of recovery capital. *Drug and Alcohol Review*. **31** (3), 334-341 (2012).
3. Laudet, A. B., White, W. L. Recovery capital as prospective predictor of sustained recovery, life satisfaction, and stress among former poly-substance users. *Substance Use and Misuse*. **43** (1), 27-54 (2008).
4. Moorhead, S. A. et al. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. *Journal of Medical Internet Research*. **15**, e85 (2013).
5. White, M., Dorman, S. M. Receiving social support online: implications for health education. *Health Education Research*. **16**, 693-707 (2001).
6. Best, D., Bliuc, A. M., Iqbal, M., Upton, K., Hodgkins, S. Mapping social identity change in online networks of addiction recovery. *Addiction Research and Theory*. **26** (3), 163-173 (2018).
7. Bliuc, A. M., Best, D., Beckwith, M., Iqbal, M. Online support communities in addiction recovery. *Addiction, behavioral change and social identity: The path to resilience and recovery*. 137, (2016).
8. Del Boca, F. K., Darkes, J. The validity of self-reports of alcohol consumption: state of the science and challenges for research. *Addiction*. **98**, 1-12 (2003).
9. Barbera, P., *Package 'Rfacebook'*. <https://cran.r-project.org/web/packages/Rfacebook/Rfacebook.pdf> (2017).
10. Wickham, H., François, R., Henry, L., Müller, K. *Package 'dplyr'*. <https://cran.r-project.org/web/packages/dplyr/dplyr.pdf> (2018).
11. Csárdi, G. *Package 'igraph'*. <https://cran.r-project.org/web/packages/igraph/igraph.pdf> (2018).
12. Walker, A., Braglia, L. *Package 'openxlsx'*. <https://cran.r-project.org/web/packages/openxlsx/openxlsx.pdf> (2018).
13. *How to get a Facebook access token which never expires*. <<https://smashballoon.com/custom-facebook-feed/access-token/>> (2018).
14. Bliuc, A. M., Best, D., Iqbal, M., Upton, K. Building addiction recovery capital through online participation in a recovery community. *Social Science and Medicine*. **193**, 110-117 (2017).
15. Best, D. et al. Overcoming alcohol and other drug addiction as a process of social identity transition: The Social Identity Model of Recovery (SIMOR). *Addiction Research and Theory*. **24**, 111-123 (2016).
16. Frings, D., Albery, I. P. The social identity model of cessation maintenance: Formulation and initial evidence. *Addictive Behaviors*. **44**, 35-42 (2015).