



Universidade Federal do Rio de Janeiro  
Instituto de Matemática  
Instituto Tércio Pacitti  
Programa de Pós-Graduação em Informática

# **Enhancing Depression Symptoms Diagnostic through Social Media Data**

**Silas Pereira Lima Filho**

**Rio de Janeiro  
2019**

Silas Pereira Lima Filho

## **Enhancing Depression Symptoms Diagnostic through Social Media Data**

**Qualificação de Doutorado** apresentada ao Programa de Pós-Graduação em Informática, Instituto de Matemática e Instituto Tércio Pacitti da Universidade Federal do Rio de Janeiro (área de concentração: Sistemas de Informação), como parte dos requisitos necessários para a obtenção do Título de Doutor em Informática.

Universidade Federal do Rio de Janeiro

Instituto de Matemática

Instituto Tércio Pacitti de Aplicações e Pesquisas Computacionais

Programa de Pós-Graduação em Informática

Supervisor: Jonice de Oliveira Sampaio, Dsc.

Rio de Janeiro

2019

---

Silas Pereira Lima Filho

EE74c

Enhancing Depression Symptoms Diagnostic  
through Social Media Data/ Silas Pereira Lima Filho. – Rio de Janeiro, 2019-  
9 f.

Supervisor: Jonice de Oliveira Sampaio, Dsc.

Tese (Doutorado) – Universidade Federal do Rio de Janeiro

Instituto de Matemática

Instituto Tércio Pacitti de Aplicações e Pesquisas Computacionais

Programa de Pós-Graduação em Informática, 2019.

1. Depression Symptoms Identification. 2. Complex Networks. 3. Social  
Network Analysis. I. Oliveira, Jonice, Advisor II. Universidade Federal do Rio de  
Janeiro III. Enhancing Depression Symptoms Diagnostic through Social Media Data

---

Silas Pereira Lima Filho

## **Enhancing Depression Symptoms Diagnostic through Social Media Data**

**Qualificação de Doutorado** apresentada ao Programa de Pós-Graduação em Informática, Instituto de Matemática e Instituto Tércio Pacitti da Universidade Federal do Rio de Janeiro (área de concentração: Sistemas de Informação), como parte dos requisitos necessários para a obtenção do Título de Doutor em Informática.

Trabalho aprovado. Rio de Janeiro, DD de MMMM de 2019:

---

**Jonice de Oliveira Sampaio, Dsc.**  
Orientador

---

**Monica Ferreira da Silva, Dsc.**  
UFRJ

# ABSTRACT

**Palavras-chaves:** Informática na Saúde Mental. Depressão. Análise de Redes Sociais.

# ABSTRACT

**Keywords:** Mental Health Informatics. Depression. Social Network Analysis.

# LIST OF FIGURES

Figure 1 – Workflow stages for the proposed methodology . . . . .	5
---	---

# LIST OF TABLES



# LIST OF ABBREVIATIONS AND ACRONYMS

ARS      *Análise de Redes Sociais*

# CONTENTS

<b>1</b>	<b>DEPRESSÃO . . . . .</b>	<b>2</b>
<b>1.1</b>	<b>Inovação . . . . .</b>	<b>2</b>
1.1.1	Gestão da Inovação . . . . .	2
<b>2</b>	<b>REVISÃO DA LITERATURA . . . . .</b>	<b>3</b>
<b>3</b>	<b>PROPOSTA DE ABORDAGEM . . . . .</b>	<b>5</b>
<b>3.1</b>	<b>Expected Contributions . . . . .</b>	<b>6</b>
<b>4</b>	<b>CONCLUSÃO . . . . .</b>	<b>7</b>
	<b>BIBLIOGRAPHY . . . . .</b>	<b>8</b>

# 1 DEPRESSÃO

## 1.1 INOVAÇÃO

lorem ipslum

### 1.1.1 Gestão da Inovação

## 2 REVISÃO DA LITERATURA

For the literature selection we have applied a systematic literature review in order to have a deeper insight from the most actual research which tackles the depression diagnosis in social media.

From a computational perspective, the problem of dealing with depression and social media can be understood as search for people who suffer the symptoms of depression. Using different methods to identify if someone has depression or not, or if some person has one of the symptoms of depression. A good amount of articles relies on natural language processing (NLP) to make a systemic analysis over the text in social media publications.

In (ZHAO; LIN; HUANG, 2018), the authors have applied text classification technique using Convolutional Neural Networks to classify depression using text analysis. The authors in (NOBLES et al., 2018) also have used neural networks in order to find patterns on periods when the risk of suicide attempt is increasing in SMS texts.

The work in (LI et al., 2016) is a qualitative study that tries to understand how is the behavior and reception of Chinese population about depression. It is a qualitative study and differs from the prior ones.

De Choudhury is an author that we would like to highlight. She is one author who has developed many articles and researches about measurement of depression on social media context. Her works vary on some types of measurement and implications of depression in some contexts of affected persons. We will cite some of these works due to the importance and relevance of her effort to contribute to this area. In (De Choudhury; COUNTS; HORVITZ, 2013), the authors start to analyze might afflicted people by depression. They have used crowdsourcing to obtain data from Twitter by people who were clinically diagnosed with depression. With this data, they have constructed a corpus and created a probabilistic model. The trained model would be relevant to indicate if a not seen post indicates depression. Similar to previous work, Tsugawa et al (TSUGAWA et al., 2015) have applied the same analysis. However the users were selected from Japan. They try to replicate the results from (De Choudhury; COUNTS; HORVITZ, 2013).

In (PARK et al., 2015), they present how activities on Facebook are associated with the depressive states of users and how depressive moods. Their goal was to raise

awareness to depression at the university where the study was conducted.

On (ANDALIBI; OZTURK; FORTE, 2017), the authors explore self-disclosures posts in Instagram. In this article, the authors use posts from people who tagged their post with #depression in order to understand what kinds of sensitive disclosures do people make on Instagram.

In (HOMAN et al., 2014), the authors bring an approach to understand what is the behaviour of users in TrevorSpace. TrevorSpace is a social media platform (social network site - SNS) where their users are people from LGBTQ. This platform aims to prevent and avoid suicide among these users community.

The author in (VEDULA; PARTHASARATHY, 2017) conduct an observational study to understand the interactions between clinically depressed users and their ego-network when contrasted with a group of users without depression. They identify relevant linguistic and emotional signals from social media exchanges to detect symptomatic cues of depression.

In (YAZDAVAR et al., 2017), authors incorporates temporal analysis of user-generated content on social media for capturing symptoms. They developed a statistical model which emulates traditional observational cohort studies conducted through online questionnaires by extracting and categorizing different symptoms of depression by modeling user-generated content in social media.

In (CHEN et al., 2018), they detected eight basic emotions and calculated the overall intensity (strength score) of the emotions extracted from all past tweets of each user. After that, created a time series for each emotion of every user to generate a selection of descriptive statistics for these time series.

### 3 PROPOSTA DE ABORDAGEM

The proposed methodology aims to combine concepts from computer science area, and from psychology topic area called psychometrics. Our methodology can be divided in three stages and it is represented in Figure 1. It has as premise a given dataset of publications in a social media platform (e.g. twitter, reddit, instagram, etc.), we should be capable to apply all the stages for any kind of social media.

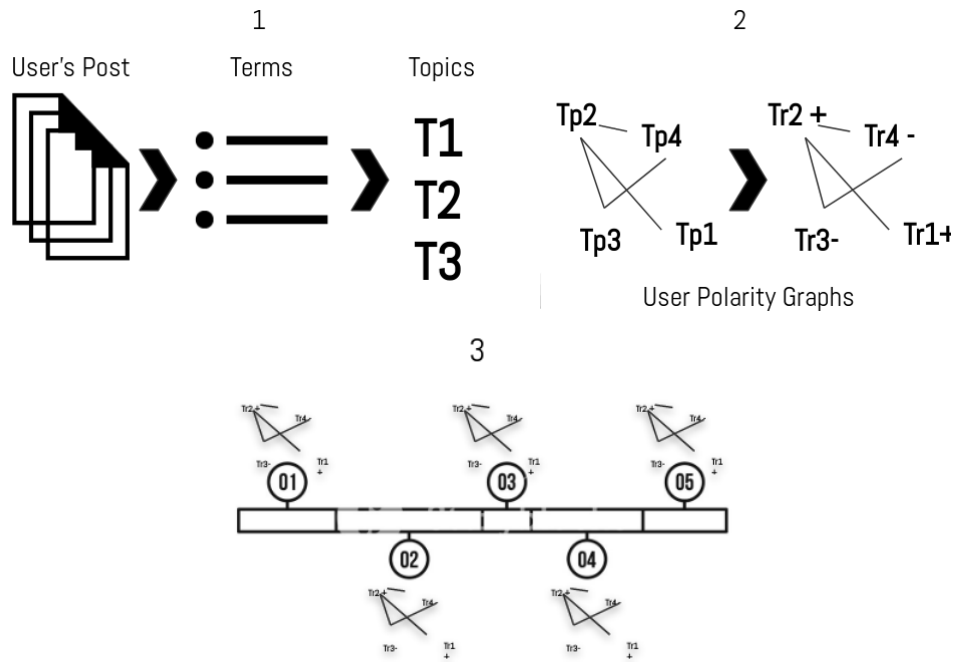


Figure 1 – Workflow stages for the proposed methodology

As most of the presented articles in Chapter 2, we also intend to apply text analysis over the datasets, specifically the technique that we want to employ is topic identification. Based on the work of (SOUSA, 2016), we aim to first identify what are the main topics of each user in a social media platform. After the topics identification, for each user, we could map the terms into a polarity graph. The polarity graph could help to identify if the most used words and terms of a user tend to be positive or negative. Related work has shown that depressive people used to manipulate more negative words. This reflects the low self-vision that this group express of themselves. With the polarity graphs of each user, in third stage we intend to analyze how these graph have evolved over the time. In this manner, we could be capable to check if someone discourse has been turned into more positive or negative. This could help to better explain the phenomena of depression inside social media. The time series analysis seems to be important, given that someone is considered depressed if he has manifested a set of symptoms for a certain

period of time. Also, with the graph of topics, we can apply social network metrics to understand how the users are connected, and how the people around them are affected by their discourse or behavior.

In the psychological approach, we intend to apply well defined questionnaires from psychometrics area on users selected from the computational approach. Psychometrics represent the theory and technique of measuring mental processes and it is applied in the fields of psychology and education. These questionnaires could be incorporated to social media users to corroborate the classification of potential users with depression.

### 3.1 EXPECTED CONTRIBUTIONS

This work still in development of methodology, with some ideas of analysis that could be made. The expected contributions are related to computer science and psychology. For the computer science, we expect that employing computational metrics like social network analysis and topic classification could support people to acquire a more proper understanding of how the phenomena of depression happen in social media, and also how the social media reflects the real life. For the health research point of view (psychology, medicine), our approach could improve how the diagnosis of depression is performed. This could aid people with few resources to enjoy a better health by the use of technology.

## 4 CONCLUSÃO

Contudo, esta pesquisa limita-se a tratar do diagnóstico e monitoramento das parcerias para auxiliar no balanceamento e manutenção da comunidade. Desta forma, o presente trabalho busca compreender as seguintes características:

- Recursos e Competências disponíveis;
- Relações de Influência e Poder;
- Relevância de Participantes;
- Confiança de Participantes

Não fazem parte do escopo deste trabalho a proposição de mecanismos de recomendação, diagnóstico de parcerias ou análise estratégica do ecossistema. Contudo, espera-se que a plataforma e o método que serão produzidos através deste projeto de pesquisa possam fomentar o desenvolvimento de outros trabalhos nesta linha.



# BIBLIOGRAPHY

ANDALIBI, N.; OZTURK, P.; FORTE, A. Sensitive Self-disclosures, Responses, and Social Support on Instagram: The Case of #Depression. In: *Proceedings of the 2017 {ACM} {Conference} on {Computer} {Supported} {Cooperative} {Work} and {Social} {Computing} - {CSCW} '17*. Portland, Oregon, USA: ACM Press, 2017. p. 1485–1500. ISBN 978-1-4503-4335-0. Disponível em: <http://dl.acm.org/citation.cfm?doid=2998181.2998243>. Citado na página 4.

CHEN, X. et al. What about Mood Swings? Identifying Depression on Twitter with Temporal Measures of Emotions. p. 8, 2018. Disponível em: <https://doi.org/10.1145/3184558.3191624>. Citado na página 4.

De Choudhury, M.; COUNTS, S.; HORVITZ, E. Social Media As a Measurement Tool of Depression in Populations. In: *Proceedings of the 5th Annual ACM Web Science Conference*. New York, NY, USA: ACM, 2013. (WebSci '13), p. 47–56. ISBN 978-1-4503-1889-1. Disponível em: <http://doi.acm.org/10.1145/2464464.2464480>. Citado na página 3.

HOMAN, C. M. et al. Social Structure and Depression in TrevorSpace. In: *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*. New York, NY, USA: ACM, 2014. (CSCW '14), p. 615–625. ISBN 978-1-4503-2540-0. Disponível em: <http://doi.acm.org/10.1145/2531602.2531704>. Citado na página 4.

LI, G. et al. SunForum: Understanding Depression in a Chinese Online Community. 2016. Disponível em: <http://dx.doi.org/10.1145/2818048.2819994>. Citado na página 3.

NOBLES, A. L. et al. Identification of Imminent Suicide Risk Among Young Adults Using Text Messages. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM, 2018. (CHI '18), p. 413:1—413:11. ISBN 978-1-4503-5620-6. Disponível em: <http://doi.acm.org/10.1145/3173574.3173987>. Citado na página 3.

PARK, S. et al. Manifestation of Depression and Loneliness on Social Networks: A Case Study of Young Adults on Facebook. In: *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. New York, NY, USA: ACM, 2015. (CSCW '15), p. 557–570. ISBN 978-1-4503-2922-4. Disponível em: <http://doi.acm.org/10.1145/2675133.2675139>. Citado na página 3.

SOUSA, D. N. F. *Automatic Research Areas Identification in C&T*. 195 p. Tese (Doutorado) — Universidade Federal do Rio de Janeiro, 2016. Citado na página 5.

TSUGAWA, S. et al. Recognizing Depression from Twitter Activity. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*, p. 3187–3196, 2015. Disponível em: <http://dl.acm.org/citation.cfm?doid=2702123.2702280>. Citado na página 3.

VEDULA, N.; PARTHASARATHY, S. Emotional and Linguistic Cues of Depression from Social Media. *Proceedings of the 2017 International Conference on Digital Health - DH '17*, p. 127–136, 2017. Disponível em: <<http://dl.acm.org/citation.cfm?doid=3079452.3079465>>. Citado na página 4.

YAZDAVAR, A. H. et al. Semi-Supervised Approach to Monitoring Clinical Depressive Symptoms in Social Media. In: *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017*. New York, NY, USA: ACM, 2017. (ASONAM '17), p. 1191–1198. ISBN 978-1-4503-4993-2. Disponível em: <<http://doi.acm.org/10.1145/3110025.3123028>>. Citado na página 4.

ZHAO, X.; LIN, S.; HUANG, Z. Text Classification of Micro-blog's "Tree Hole" Based on Convolutional Neural Network. In: *Proceedings of the 2018 International Conference on Algorithms, Computing and Artificial Intelligence*. New York, NY, USA: ACM, 2018. (ACAI 2018), p. 61:1—61:5. ISBN 978-1-4503-6625-0. Disponível em: <<http://doi.acm.org/10.1145/3302425.3302501>>. Citado na página 3.