Social Media Influence Analyzer

Mohammed Z. Guniem  
Department of Computer and Electrical Technology  
*University of Stavanger*Stavanger, Norway  
mghunime@yahoo.no

*Abstract*—This electronic document is a “live” template and already defines the components of your paper [title, text, heads, etc.] in its style sheet. *\*CRITICAL: Do Not Use Symbols, Special Characters, Footnotes, or Math in Paper Title or Abstract*. (*Abstract*)

Keywords—component, formatting, style, styling, insert (key words)

# Introduction

Upon the rise of the still ongoing digital revolution, many digital platforms had a series of success stories in transforming the way society members interacts with each other on the daily basis. People around the world are no longer limited by distance or time to be able to reach out for one another and participate in social activities or events. Because of this increase of socializing on multiple popular digital platforms, a big amount of digital data is exchanged and recorded about each activity a certain user performs on those digital platforms.

Such data has a great potential in revealing the strengths of influences between network users, just like in real life every action a user commit and how society members react to this action can serve as an object for analysis which helps in drawing a big but rather detailed picture of how users influence each other across many societies and fields.

The aim of this research is establishing a ground foundation for extracting information about user activities on social media and use such information to detect social influence between network users. Such foundation is desired to make up the core of a future technical solution that enables social media analyzers with little or no technical experience in data processing and visualization to perform social media analysis on regular periods with a continuous timeline.

To serve this purpose, we start by determining the common characteristics in available functionalities between the most popular social media platforms, then produce a data model based on similarity between them. This data model is crucial to guarantee the reliability and flexibility of algorithms and technical systems that builds on top of this research; both reliability and flexibility are two sides of the same coin, as when basing the data model of the system on similarities between data provided by as many social media platforms as possible, we naturally widen the range of analysis potentials of any applying digital platform, making the solution more flexible and highly reliable and adaptable for social media analyzers.

After establishing an agreement on the data model to be used for collecting and storing crawled data from social media, we dive into the main core functionality of detecting social influence between network users. Multiple techniques of detecting such influences will be implemented to fit the different needs and purposes of the final application of analysis, regardless if the purposes are independent or can be tighten together in combinations. The desired result is a user influence graph where each node represents a participating user, while each edge between two given users representing the influence between these two users with respect to direction of influence and holding the strength and area of the influence whether it is in sport, politic, economy etc.

Following the previous effort, we evaluate the performance of the user graph model and go through test results from both dummy and real-life data using crawled data from a rising social media platform called “Reddit”. We will try to highlight the most interesting and useful features of the produced user graph and push its power of detecting influencers and their area of influence to the limit. The final test and evaluation process is a vital and necessary step to rely on the quality of the produced user graph model in any future analysis and technical solution.

This was a brief introduction of the upcoming research in a nutshell, but first let’s go through some interesting attempts in revealing user influence from social media.

# Related Works

Among the community of computer science, a wide variety of studies has focused on extracting information from the available social media platforms, and a big amount of effort has been dedicated to reveal influences between users to better understand the behaviour of individuals for many both commercial and non-commercial purposes. Research of social influence takes different forms and vary in size and scope, while some researchers take on the very fundamentals of detecting social influence others dive through to reveal specific influences and hidden behaviour patterns on different levels.

## Detecting and measuring user influence in social media

A good fundamental approach is described by a social network analysis carried out by Y. Guo, J. Cao & W. Lin. The fellow researchers are dividing the influence evaluation models into 2 main categories; the first category is based on network topology which measure social influence of a certain user by considering his degree, shortest path, and random walk characteristics, while the second category bases the influence between users on their interactions through different activities organized in tree-like structures like submissions and multilevel comments. However, and despite the reasonably good classification and overview these researchers offers, their paper lacks some proven results of an experimental approach [1].

## Data-driven Influence Learning

Identify applicable funding agency here. If none, delete this text box.

A short but rather interesting experimental and mathematical approach is introduced by a paper on Data-driven Influence Learning in Social Networks published by F. Wang, W. Jiang, G. Wang & D. Xie. In this paper, the process of influence diffusion is divided into two parts: the launcher (influence strength) and the receiver (influence threshold) which can generate an accurate and finer grained influence diffusion model according to this research. Furthermore, the researchers highlight the importance of having a solid criterium when scoring the strength and threshold properties of social influences. Another important acknowledgment is the difficulty and complexity associated with detecting influence relationships between users as a by-product of big datasets that usually include a considerable amount of noisy or less important datapoints, making it essential for any algorithm used in learning and testing the influence models to perform a minimal scan over the data in the most efficient way possible. [2]

## Gathering data and Crawling alternatives

Most well-known providers of social media platforms offer

developers and data scientist multiple endpoints and ways to extract data from their platforms for development or analysis. Some research spots the light on this initial aspect of gathering data from social media platforms. One significant research is one that mainly describe the alternative of Pushshift Reddit Dataset by J. Baumgartner, S. Zannettou, B. Keegan, M. Squire and J. Blackburn. [3] The research paper offers an undirected but also claimed to be a more efficient and flexible way to gather data from the “Reddit” social platform than by using the official Reddit API endpoint.

It also gives an excellent brief description of the FAIR data principles which is highly relevant when choosing the source of data especially when it comes to accessibility and findability.

Another advantage of this research is its extension in discussing a series of the other major alternatives for gathering data from Reddit, highlighting their strengths and weaknesses in a constructive manner.

## Classification of topics in social media platforms

As mentioned in the introduction, we are set to determine

the category of a detected influence between users, this opens up for the use of artificial intelligence for the purpose of classification between different topics where a certain user activity might fit in. In a research about annotating and detecting topics in social media forum and modelling the annotation to derive directions carried out by B. Athira, J. Jones, S. M. Idicula, A. Kulanthaivel and E. Zhang, a practical case study from an online health community was represented to give a good introduction of data pre-processing and cleaning, preceding to construct a reasonable mathematical approach for training and testing of a produced classification machine learning model.

Furthermore, the research offers a solution for a much-needed ability to minimize the amount of training data and dealing with the negative effects of label imbalance in a training dataset.

Another contribution of this research is the use of various deep learning algorithms to classify posted content such as CNN, LSTM and BiLSTM, all in which enable the researchers to achieve a promising F1-score of about 0.75 to 0.80 in topic classification.

The research has an excellent and well-performed evaluation and testing phase, where metrics of evaluation are carefully examined and explained in a good scientific approach [4].

## Choosing a study case social media platform

Determining which social media platform to crawl for testing and evaluation purposes is an important choice in the path to producing a user influence model that is flexible and useable in as many social media platforms as possible, this is why it is desirable to base the real life study case on actual real life datasets from a digital media platform that shares common user functionalities with as many popular social media platforms as possible, examples of such functionalities are post, submissions, comments, and upvotes or commonly known as likes.

A social media platform that satisfies all these user functionalities is “Reddit” which is examined by the research called “Information and Social Analysis” carried out by T. Steinbauer at the University of California, Santa Barbara. [5] Steinbauer starts off with a brief but very constructive comparison between the most popular social news sites with Reddit included. The core of Steinbauer’s research lays in his analysis of subreddits, submissions and comments on Reddit which can serve as a foundation prove for why Reddit should be used in evaluating the performance of a user influence graph model and its ability to view the most influencing users in a social media platform, the reason behind this is that Steinbauer gives a detailed analysis on which subreddits seems to have the most of activity and also precedes in building into constructing a user graph model that helps showing which user has the highest influence based on users interactions through comments.

However, submission authors are not included in the dataset of the constructed user graph, making the user graph less reliable when ignoring the often-significant role of posters in generating discussions on social media. Another downside of Steinbauer’s user graph model is not using any other criteria than user interaction through comments, such as the upvote score or number of thread comments to a certain comment or submission.

Although Steinbauer has introduced a detailed overview of his evaluations and analysis’s results, there is still a big question mark on the technical details as no algorithms for constructing the user graph has been presented in detail.

Last checkpoint, next start writing about your solution

# Prepare Your Paper Before Styling

Before you begin to format your paper, first write and save the content as a separate text file. Complete all content and organizational editing before formatting. Please note sections A-D below for more information on proofreading, spelling and grammar.

Keep your text and graphic files separate until after the text has been formatted and styled. Do not use hard tabs, and limit use of hard returns to only one return at the end of a paragraph. Do not add any kind of pagination anywhere in the paper. Do not number text heads-the template will do that for you.

## Abbreviations and Acronyms

Define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract. Abbreviations such as IEEE, SI, MKS, CGS, sc, dc, and rms do not have to be defined. Do not use abbreviations in the title or heads unless they are unavoidable.

## Units

* Use either SI (MKS) or CGS as primary units. (SI units are encouraged.) English units may be used as secondary units (in parentheses). An exception would be the use of English units as identifiers in trade, such as “3.5-inch disk drive”.
* Avoid combining SI and CGS units, such as current in amperes and magnetic field in oersteds. This often leads to confusion because equations do not balance dimensionally. If you must use mixed units, clearly state the units for each quantity that you use in an equation.
* Do not mix complete spellings and abbreviations of units: “Wb/m2” or “webers per square meter”, not “webers/m2”. Spell out units when they appear in text: “. . . a few henries”, not “. . . a few H”.
* Use a zero before decimal points: “0.25”, not “.25”. Use “cm3”, not “cc”. (*bullet list*)

## Equations

The equations are an exception to the prescribed specifications of this template. You will need to determine whether or not your equation should be typed using either the Times New Roman or the Symbol font (please no other font). To create multileveled equations, it may be necessary to treat the equation as a graphic and insert it into the text after your paper is styled.

Number equations consecutively. Equation numbers, within parentheses, are to position flush right, as in (1), using a right tab stop. To make your equations more compact, you may use the solidus ( / ), the exp function, or appropriate exponents. Italicize Roman symbols for quantities and variables, but not Greek symbols. Use a long dash rather than a hyphen for a minus sign. Punctuate equations with commas or periods when they are part of a sentence, as in:

*a**b* 

Note that the equation is centered using a center tab stop. Be sure that the symbols in your equation have been defined before or immediately following the equation. Use “(1)”, not “Eq. (1)” or “equation (1)”, except at the beginning of a sentence: “Equation (1) is . . .”

## Some Common Mistakes

* The word “data” is plural, not singular.
* The subscript for the permeability of vacuum **0, and other common scientific constants, is zero with subscript formatting, not a lowercase letter “o”.
* In American English, commas, semicolons, periods, question and exclamation marks are located within quotation marks only when a complete thought or name is cited, such as a title or full quotation. When quotation marks are used, instead of a bold or italic typeface, to highlight a word or phrase, punctuation should appear outside of the quotation marks. A parenthetical phrase or statement at the end of a sentence is punctuated outside of the closing parenthesis (like this). (A parenthetical sentence is punctuated within the parentheses.)
* A graph within a graph is an “inset”, not an “insert”. The word alternatively is preferred to the word “alternately” (unless you really mean something that alternates).
* Do not use the word “essentially” to mean “approximately” or “effectively”.
* In your paper title, if the words “that uses” can accurately replace the word “using”, capitalize the “u”; if not, keep using lower-cased.
* Be aware of the different meanings of the homophones “affect” and “effect”, “complement” and “compliment”, “discreet” and “discrete”, “principal” and “principle”.
* Do not confuse “imply” and “infer”.
* The prefix “non” is not a word; it should be joined to the word it modifies, usually without a hyphen.
* There is no period after the “et” in the Latin abbreviation “et al.”.
* The abbreviation “i.e.” means “that is”, and the abbreviation “e.g.” means “for example”.

An excellent style manual for science writers is [7].

# Using the Template

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your conference for the name of your paper. In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper; use the scroll down window on the left of the MS Word Formatting toolbar.

## Authors and Affiliations

**The template is designed for, but not limited to, six authors.** A minimum of one author is required for all conference articles. Author names should be listed starting from left to right and then moving down to the next line. This is the author sequence that will be used in future citations and by indexing services. Names should not be listed in columns nor group by affiliation. Please keep your affiliations as succinct as possible (for example, do not differentiate among departments of the same organization).

### For papers with more than six authors: Add author names horizontally, moving to a third row if needed for more than 8 authors.

### For papers with less than six authors: To change the default, adjust the template as follows.

#### Selection: Highlight all author and affiliation lines.

#### Change number of columns: Select the Columns icon from the MS Word Standard toolbar and then select the correct number of columns from the selection palette.

#### Deletion: Delete the author and affiliation lines for the extra authors.

## Identify the Headings

Headings, or heads, are organizational devices that guide the reader through your paper. There are two types: component heads and text heads.

Component heads identify the different components of your paper and are not topically subordinate to each other. Examples include Acknowledgments and References and, for these, the correct style to use is “Heading 5”. Use “figure caption” for your Figure captions, and “table head” for your table title. Run-in heads, such as “Abstract”, will require you to apply a style (in this case, italic) in addition to the style provided by the drop down menu to differentiate the head from the text.

Text heads organize the topics on a relational, hierarchical basis. For example, the paper title is the primary text head because all subsequent material relates and elaborates on this one topic. If there are two or more sub-topics, the next level head (uppercase Roman numerals) should be used and, conversely, if there are not at least two sub-topics, then no subheads should be introduced. Styles named “Heading 1”, “Heading 2”, “Heading 3”, and “Heading 4” are prescribed.

## Figures and Tables

#### Positioning Figures and Tables: Place figures and tables at the top and bottom of columns. Avoid placing them in the middle of columns. Large figures and tables may span across both columns. Figure captions should be below the figures; table heads should appear above the tables. Insert figures and tables after they are cited in the text. Use the abbreviation “Fig. 1”, even at the beginning of a sentence.

1. Table Type Styles

| Table Head | Table Column Head | | |
| --- | --- | --- | --- |
| Table column subhead | Subhead | Subhead |
| copy | More table copya |  |  |

1. Sample of a Table footnote. (*Table footnote*)
2. Example of a figure caption. (*figure caption*)

Figure Labels: Use 8 point Times New Roman for Figure labels. Use words rather than symbols or abbreviations when writing Figure axis labels to avoid confusing the reader. As an example, write the quantity “Magnetization”, or “Magnetization, M”, not just “M”. If including units in the label, present them within parentheses. Do not label axes only with units. In the example, write “Magnetization (A/m)” or “Magnetization {A[m(1)]}”, not just “A/m”. Do not label axes with a ratio of quantities and units. For example, write “Temperature (K)”, not “Temperature/K”.

##### Acknowledgment *(Heading 5)*

The preferred spelling of the word “acknowledgment” in America is without an “e” after the “g”. Avoid the stilted expression “one of us (R. B. G.) thanks ...”. Instead, try “R. B. G. thanks...”. Put sponsor acknowledgments in the unnumbered footnote on the first page.

##### References

[1] Y. Guo, J. Cao and W. Lin, "Social Network Influence Analysis," 2019 6th International Conference on Dependable Systems and Their Applications (DSA), 2020, pp. 517-518, doi: 10.1109/DSA.2019.00093. J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[2] F. Wang, W. Jiang, G. Wang and D. Xie, "Data-Driven Influence Learning in Social Networks," 2017 IEEE International Symposium on Parallel and Distributed Processing with Applications and 2017 IEEE International Conference on Ubiquitous Computing and Communications (ISPA/IUCC), 2017, pp. 1179-1185, doi: 10.1109/ISPA/IUCC.2017.00177.

[3] Baumgartner, Jason & Zannettou, Savvas & Keegan, Brian & Squire, Megan & Blackburn, Jeremy. (2020). The Pushshift Reddit Dataset.

[4] Balakrishnan, Athira & Jones, Josette & Idicula, Sumam & Kulanthaivel, Anand & Zhang, Enming. (2021). Annotating and detecting topics in social media forum and modelling the annotation to derive directions-a case study. Journal of Big Data. 8. 10.1186/s40537-021-00429-7.

[5] T. Steinbaur, ‘‘Information and social analysis of Reddit,’’ inProc.TROYSTEINBAUER CS. UCSB. EDU, 2012, pp. 1–12. [Online]. Available:http://snap.stanford.edu/class/cs224w-2011/proj/tbower\_Finalwriteup\_v1.pdf

**IEEE conference templates contain guidance text for composing and formatting conference papers. Please ensure that all template text is removed from your conference paper prior to submission to the conference. Failure to remove template text from your paper may result in your paper not being published.**