

Survey Paper On Sock Puppet Detection

Abhijnan Chakraborty¹

Vaibhav Saha²

Harshit Kumar³

Rishabh Verma⁴

¹Department of Computer Science, IIT Delhi

²Department of Computer Science, IIT Delhi

³Department of Computer Science, IIT Delhi

⁴Department of Computer Science, IIT Delhi

, , , cs5200123@iitd.ac.in

Abstract -

This document sets out the requirements for preparing manuscripts for the ISARC. It is essential that all manuscripts conform to these instructions. Note that this document contains both general instructions and structures specific to the preparation of manuscripts using L^AT_EX. For instructions on how to prepare manuscripts using MS Word, please use the MS Word template.

Keywords -

Instructions; Formatting; Authors; ISARC

1 Introduction

This L^AT_EX template and the instructions it contains will enable you to prepare your manuscript in an electronic format (pdf), ready for submission and peer review. It is therefore essential that these instructions be carefully followed.

Your manuscript must be submitted to our online peer-review management system.

The recommended length for Scientific Papers is 6 to 8 pages (4 pages for Short Papers), including text, figures, tables, and references. For more information about the paper types, refer to the “Paper and Presentation Types” on the IAARC website. Electronic files should not exceed 5 megabytes.

Authors are responsible for ensuring the accuracy of all information contained in their manuscripts (e.g., proper names of organizations, data and findings, references, etc.). Manuscripts must be prepared in proper English. SI units or acceptable metric equivalents must be used throughout.

2 Related Work

In the context of Sock puppet detection, Zaher Yamak et al.[7] has used some notable features which

has contributed their work in the field of sock puppet detection. The dataset used by this paper consists of wikipedia dataset. Dataset extraction consists of 10 TB uncompressed which is compressed to 100 GB by 7-Zip. From this, 118414 sock puppet accounts got filtered. After this, in the account selection, 12088 groups were identified which contains 2-557 members. There is a random selection of 5000 sock puppets who are in a group of more than 3 sockpuppets. They are mixed with 5000 random active accounts. After this there is feature selection process, in the paper the author says that by using these features, the accuracy achieved is 99.8%. Some notable features are:-

- **The number of user’s contributions by namespaces:** This contains the categorization of user contribution into 6 categories: article, article discussion, user page, user discussion page, project namespace. This categorization strategy offers insights into user focus areas and communication patterns, enhancing the comprehension of their impact on content and community dynamics.
- **The frequency of revert after each contribution in the articles:**
- **The average of bytes added and removed from each revision:** The hypothesis is that the manipulation goal of the Sockpuppeteer can be extracted from its addition/removal behavior. The manipulation goal of a Sockpuppeteer may be either to add and publish a particular piece of (mis) information or to remove (part of) a previous contribution.
- **The average of contribution in the same article:**
- **The interval between the user’s registration and his first contribution:** The underlying assumption

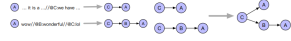


Figure 1. Propagation Tree Construction (Reference:[11])

is that a manipulator creates at the beginning of its manipulation attempt many accounts, and then leaves them sleeping to use them separately as backup when an active account is blocked.

After this, they do a 10-fold cross-validation technique to assess different models. The algorithm and metrics used are TPR, FPR, F-measure, precision, and MCC. The experimental results are compared with some previous works one is Gao et. al who used the verbal attributes as features which got about 93.9% accuracy, after this Solorio paper which was based on textual features, alphabet count, number of tokens and use of words without vowels have accuracy of about 68.83% using SVM. Also, this is not so accurate if user changes his writing style. Another paper of Yang that uses non verbal attributes are invitation frequency, outgoing request accept, incoming request are taken into account which gives an accuracy of about 71.3% using SVM classifier. The paper concludes that his technique exceeds many previous algorithms and plan to explore other social media to verify the feature set in other settings, such as forums or twitter.

Jiacheng et al.[11] discusses how its previously ignored a very crucial property present in the sockpuppet behavior, the propagation characteristics. The paper uses the property that the propagation tree of a sockpuppet is deeper and highlights that the message is reposted by sockpuppet will be spread far. The paper talks about the **Difference of pairwise accounts** which is used in the paper as a useful property, shows the sockpuppet pair is more similar than others through three dimensions: size, depth, and width. It is reasonable that the pairwise sockpuppets behave similarly.

The propagation tree is constructed by following as shown in the figure 1, Each reposting log will represents an information propagation process, such as "wow!!!B:wonderful!!@C:lol". Based on the practice of refereeing to another account in a tweet via "@username" convention, we extract the usernames from reposting log and construct the propagation trees to represent the information propagation process of an account.

The paper deals the whole problem of sock puppet detection using the graph terminologies in which assume $G = (V, E)$ be a social network, where V is a set of accounts,

$E \in V \times V$ is a set of repost relationship, and $e_{vu}^i \in E$ denotes repost relationship of message i between account v and $u(v, u \in V)$ which reflects propagation of information over G . In this suppose u represents a user so the sock puppet account detection goes through some notable features such as:-

•

2.1 Page Size and Layout

Your manuscript must be prepared for A4-size (210 x 297 mm) paper. Use the margin settings specified in Table 1 and do not number the pages of the paper. The L^AT_EX template already implements those margins.

Table 1. Manuscript margins	
Margin	A4 (210 x 297 mm)
Top	3.5 cm
Bottom	3.5 cm
Left	2.5 cm
Right	2.0 cm

2.1.1 Headers and footers

Do not add anything in the footers or headers, even page numbers. These will be added by the IAARC Technical Committee when compiling the Proceedings.

2.2 Type Style and Spacing

All text must use the Times New Roman font. This L^AT_EX template already implements this.

2.2.1 Title Page

The title page of your manuscript is illustrated by the first page of these instructions. The title page must not exceed one (1) page in length. It contains:

1. The title of the paper in bold 18 points Times New Roman, centered with first letters in capitals.
2. The authors' names, separated from the title by a one blank line, are centered and in size 11, bold Times New Roman.
3. The authors' affiliations and addresses are put immediately below the names, centered and single-spaced, in size 10.
4. Email addresses are inserted below the affiliations, also in size 10.

This template already implements all these requirements. Simply use the commands `\title{}`, `\author{}`, `\affiliation{}` and `\email{}` as illustrated here.

2.2.2 Body of paper

The body of the paper follows the front matter. It contains two columns (with 0.5cm separation) and begins with:

- The major heading “Abstract –”, in bold.
- The body of the abstract, not to exceed 250 words in length, in bold Times New Roman, fully justified, the first line is indented.
- The major heading “Keywords –” in bold, separated from the last line of the abstract by one blank line.
- The list of keywords, not to exceed ten (10) words, in bold Times New Roman, left justified, indented, and separated by commas. Please add those keywords that you would use if you were searching for your paper.

The main text of the paper begins after. Separate sections of the main text in accordance with the Headings guidelines below

This template already implements all these requirements. Simply use the environments `abstract` and `keywords` as illustrated in this template.

2.2.3 Headings

All headings must be in black and in bold face. The manuscript will typically have maximum three levels of headings maximum: Level 1, Level 2, and Level 3. Level-1 headings (e.g. Introduction, Background, Discussion, Conclusions, Acknowledgments, References) have font size 12; Level-2 headings have font size 11; and Level-3 headings have font size 10.

Words in the headings should be all lower case, except the first letter of the first word which must start with a capital letter. The headings are positioned at the left margin. They are numbered in the form “1.” for Level-1 headings, “1.1.” for Level-2 headings, and “1.1.1” for Level-3 headings.

The commands `\section{}`, `\subsection{}` and `\subsubsection{}` employed in this template implement the requirements above for headings Level 1, 2 and 3. Note that you may also use the command `\paragraph{}`, if this suits better your needs.

2.2.4 Text

Text paragraphs are single-spaced and fully justified, with the first line indented 0.5 cm. Do not use blank lines between paragraphs unless you feel it important to really highlight a change of topic (in such a case, you should also consider using sub-headings).

Note that, to start a new paragraph in \LaTeX , simply leave a blankline in the `.tex` file.

2.2.5 Lists

1. Numbered lists should be presented using the environment `enumerate` like in this example. This environment applies the numbering and defines the format and spacing automatically.
 2. Add the option `[noitemsep]` to the environment to have no additional space between numbered list items.
- Bulleted lists should be presented using the environment `itemize` like in this example. This environment applies the numbering and defines the format and spacing automatically.
 - Add the option `[noitemsep]` to the environment to have no additional space between bulleted list items.

Note that the two list above were generated using the environments and options described within them.

2.2.6 Footnotes

Do not use footnotes. Incorporate all required information in the body of the paper.

2.2.7 Equations and symbols

Simple mathematical expressions and sub- and superscripted characters, such as SO_4^2 , are inserted in the text. Do not embed equations as an image. \LaTeX is valued particularly for its performance in handling equations and symbols.

Equations are placed on separate lines, centered and numbered consecutively in parentheses at the right-hand margin. A blank line precedes and follows each equation.

To achieve this, simply use the environment ‘*equation*’ as in the examples below.

$$ZnS + 3/2O_2 \rightarrow ZnO + SO_2 \quad (1)$$

$$ZnS + 3/2O_2 = ZnO + SO_2 \quad (2)$$

$$E = 1.23 - 0.06pH \quad (3)$$

The nomenclature and units for symbols must be defined in the text. Alternatively, where the number of symbols is large, a special Section **Nomenclature** should be used at the end of the manuscript.

2.3 Figures

Figures should appear as close as possible to their first citation. They should be numbered consecutively using Arabic numerals and their title should be centered below the figure.

Figures, such as graphs and diagrams, should be embedded in vector format (e.g. `.eps`, `.pdf` or `.emf`), if at

all possible. Otherwise, figures such as pictures should be with high resolution (300 dpi) when published at 100% (e.g., images at 72 dpi are in reality 25% of the required resolution).

For example, Figure 1 in this manuscript is embedded as a picture with high resolution (but, ideally, such graph should be prepared and embedded in vector format). Lines and lettering must be large enough (minimum 0.35pt thickness) to remain clearly legible when printed at 100%. For maps and similar figures, be sure to place a scale marker on the picture or photograph. Color is preferable; any grayscale figure requires sharp contrast. Please, refrain from using frames around figures, or using shaded backgrounds. Figures can significantly increase the size of the electronic file. Excessively large files tend to complicate and slow down the editing process. Therefore, you must make every effort to reduce the size of the electronic files of your figures. Embedding figures in vector format, as mentioned above, can significantly reduce file size while simultaneously enhancing figure quality.

The caption should be formatted with the word ‘Figure’ followed by the figure number, a period, a space, and the title. Each figure and caption should be separated from the adjacent text with one blank line. To achieve all this formatting, simply use the environment `figure` as in Figure 2.

A figure too wide to fit in one column may be placed across the width of the page (i.e. between the margins) by using the environment `figure*`. A figure too wide to fit between the margins may be placed in landscape orientation (sideways format), on a page by itself, with the bottom of the figure to the right of the page. To achieve this, embed the environment `figure` within an environment `landscape`.

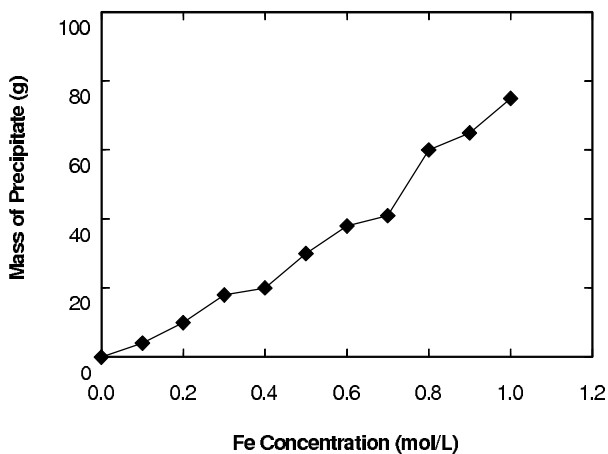


Figure 2. Effect of iron concentration on the amount of precipitate formed during hydrolytic precipitation from waste processing solutions

2.4 Tables

Tables should appear as close as possible to their first citation. They should be numbered consecutively using Arabic numerals and their title should be centered above the table.

The caption should be formatted with the word ‘Table’ followed by a character space, the table number, a period, another character space, and the title. Separate each table from the adjacent text with one blank line. See Table 2 for example.

Table-wide lines (horizontal 0.5 point thickness) separate the title from the column headings, the column headings from the body of the table, and the table from the following text. Avoid vertical lines and avoid the use of horizontal lines between the various rows of data.

Text in the tables should have the same size as the main text, but not be indented and with space before and after it. Table 2 shows a good example. A table too wide to fit in one column may be placed across the width of the page (i.e. between the margins) by using the environment `table*`. A table too wide to fit between the margins may be placed in landscape orientation (sideways format), on a page by itself, with the bottom of the figure to the right of the page. To achieve this, embed the environment `table` within an environment `landscape`.

Table 2. Electron microprobe analyses of sphalerite grains in the Kidd Creek “C” concentrate

Element	Average(wt %)	Range(wt%)
Zn	60.8	59.6 - 63.3
Fe	5.82	3.54 - 6.95
Cd	0.30	0.12 - 0.42
S	3.31	33.6 - 33.5

2.5 Cross-referencing headings, numbered equations, figures and tables

Refer to a figure as ‘Figure X’, not using its relative position. For achieving this effectively, use the command `\label{}` within the environment `figure` to label each figure, and then use the command `\ref{}` in the text to refer to the label.

The same guidance applies to Tables and Equations, as well as Headings. In the case of headings, the command `\label{}` must be declared right after the heading commands (e.g. after the command `\section{}`).

Note that all the figures, tables, equations and the main headings in this template all have a label and are referred to in the text using the command `\ref{}`.

2.6 References

References to the literature should be cited in the main text with an Arabic number in square brackets [1]. References should then appear in cited order at the very end of your paper in the Section **References**. Start each reference on a new line with its number in square brackets [1]. Citation formats are given below for: a journal article [1], conference paper [2], book [3], and a website [4].

The formatting of the references is done automatically by the L^AT_EX template. All you need to do is add your references in the ISARC.bib Bibtex file and cite the papers in the text using the command `\cite{}` (as well as `\citet{}`).

Within the Bibtex file, please ensure that all references that have a DOI actually include it. It will then appear in the compiled references.

References

- [1] A. ASurname, B. BSurname, and C. CSurname. A journal article title. *The Journal*, 1(2):123–345, 2014. doi:10.22260/ISARC2017/0001.
- [2] D. DSurname and E. ESurname. A conference paper title. In *Proceedings of the conference*, pages 12–34, Sydney, Australia, 2014.
- [3] F. FSurname. *A book title*, volume 2(3). The Publisher, Publication address, 2014.
- [4] G. GSurname. Title of the web content. On-line: <http://abcd.efg.hij>, Accessed: 01/04/2019.