Information Exposure Maximization Problem with Heuristic and Evolutionary Algorithms

Layheng Hok   
*Department of Computer Science and Engineering  
Southern University of Science and Technology*Shenzhen, China  
12210736@mail.sustech.edu.cn

# Introduction

This template, modified in MS Word 2007 and saved as a “Word 97-2003 Document” for the PC, provides authors with most of the formatting specifications needed for preparing electronic versions of their papers. All standard paper components have been specified for three reasons: (1) ease of use when formatting individual papers, (2) automatic compliance to electronic requirements that facilitate the concurrent or later production of electronic products, and (3) conformity of style throughout a conference proceedings. Margins, column widths, line spacing, and type styles are built-in; examples of the type styles are provided throughout this document and are identified in italic type, within parentheses, following the example. Some components, such as multi-leveled equations, graphics, and tables are not prescribed, although the various table text styles are provided. The formatter will need to create these components, incorporating the applicable criteria that follow.

# Preliminary

The IEMP is framed as an optimization algorithmic problem aimed at reducing the effects of echo chambers and filter bubbles, where users are frequently limited to similar viewpoints, isolating them within their own information spheres. This study examines a social network represented as a graph *G = (V, E)* where *V* represents users (nodes) and *E* represents social connections (edges) between them. The objective in the IEMP is to identify two sets of users that achieve the highest balanced exposure to diverse information.

The following sections provide essential definitions of terminology and notations used, and a formal definition of the problem.

## Terminology and Notations

* **Social Network** - *G =* (*V, E*) represents the network graph, where *V* = {𝑣*1*, 𝑣*2*, …, 𝑣*n*} represents the set of nodes, and *E* ⊆ *V*×*V* represents the edges between nodes.
* **Campaigns** – *C =* {*c1*, *c2*} represents two campaigns; each campaign holds its own viewpoint.
* **Initial Seed Set** – *Ii* ⊆ *V*, *i* ∈ {1, 2} represents the initial seed set for campaign *ci* .
* **Balanced Seed Set** – *Si* ⊆ *V*, *i* ∈ {1, 2} represents the target seed set needed to find for each campaign *ci*.
* **Budget** – *k* represents the size’s upper bound of the two balanced seed sets; that is, *|S1| + |S2| ≤ k*.
* **Diffusion Probability** – *Pi =* {*pi*(u, v)∣(*u*, *v*)∈ *E*}, *i* ∈ {1, 2} represents the edge weight associated with campaign *ci* where *pi*(u, v) represents the probability of node *u* activating node *v* under each campaign *ci*.
* **Diffusion Model** – *M* represents the stochastic process for the seed *Ui = Ii*∪ *Si*, which initiates the spread of information on graph *G*. It is assumed that information from the two campaigns spreads through the network following the independent cascade (IC) model, with each campaign’s message spreading independently-this is known as heterogeneous propagation. The diffusion process for the first campaign (with the second campaign following the same process) proceeds as follows:
  1. At step *t* = 0, nodes in the seed set *U1*are activated, while all the other nodes remain inactive.
  2. In each step *t*, any active user *u* for campaign *c1* attempts to activate each of its inactive outgoing neighbors *v* with a probability *p1*(u, v). This activation is like flipping a coin with a probability of heads equal to *p1*(u, v): if heads, *v* is activated; if tails, *v* remains inactive. Each active user *u* has only one chance to activate each neighbor for campaign *c1*​, after which *u* stays active but ceases further activation efforts.
  3. The diffusion process ends when no more nodes can be activated.
* **Exposed Node Set** – *ri*(*U*) represents the set of vertices that can be reached from *U* through the cascade process mentioned above for campaign *ci*, for a given seed set *U*. Note that in a single propagation step, *ri*(*U*) includes not only nodes that were successfully activated by *U* but also nodes that were targeted for activation by *U* but remained inactive. Since the diffusion process is stochastic, *ri*(*U*) is a random variable.

## Problem Formulation

Given a social network *G =* (*V, E*), two sets of *I1* and *I2* of the initial seed sets for the two campaigns, and a budget *k*, the goal of the IEMP is to identify two sets *S1* and *S2*​ such that *|S1| + |S2| ≤ k* and to maximize the expected number of vertices that are either influenced by both campaigns or remain unaffected by both; that is,

max Φ(*S1*, *S1*) = max 𝔼[|𝑉 ∖ (𝑟*1* (*I1*∪ *S1*) △ 𝑟*2* (*I2*∪ *S2*))|]

such that *|S1| + |S2| ≤ k* and *S1*, *S2* ∈ *V*

# Methodology

As introduced earlier, IEMP’s diffusion model is based on a stochastic process. The two algorithms outlined below will leverage this diffusion model within their respective procedures. To facilitate this, it is essential to define an algorithm that accurately represents the diffusion process, which can be accomplished through the application of a breadth-first search (BFS) approach.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Algorithm 1: Influence Diffusion Via BFS** | | | | | | |
|  | **Input**: *graph*(*G*), *seed* (*Ui = Ii*∪ *Si*), *campaign* (campaign index) | | | | | |
|  | **Output**: *active* (set of active nodes), *exposed* (set of exposed nodes) | | | | | |
| 1 | Initialize *queue* as an empty queue | | | | | |
| 2 | Initialize *active* as an empty set | | | | | |
| 3 | Initialize *exposed* as an empty set | | | | | |
| 4 | **for** *node* in seed **do** | | | | | |
| 5 |  | *queue*.Add(*node*) | | | | |
| 6 |  | *active*.Add(*node*) | | | | |
| 7 |  | *exposed*.Add(*node*) | | | | |
| 8 | **while** *queue* is not emoty **do** | | | | | |
| 9 |  | *node* ← *queue*.Remove() | | | | |
| 10 |  | **for** *neighbo*r, *probab* in *graph*.GetAdjLis*t*(*node*) **do** | | | |
| 11 |  |  | **if** not *exposed*.Contains(*neighbor*) **then** | | |
| 12 |  |  |  | *exposed*.Add(*neighbor*) | |
| 13 |  |  | **if** not *active*.Contains(*neighbor*) **then** | | |
| 14 |  |  |  | **if** *probab* ≥ RandomValue() **then** | |
| 15 |  |  |  |  | *active*.Add(*neighbor*) |
| 16 |  |  |  |  | *queue*.Add(*neighbor*) |
| 17 | **return** *active*, *exposed* | | | | | |

## Heuristic Search

This section outlines a heuristic approach for solving the IEMP using a Monte Carlo-based greedy selection strategy.

The main idea of this heuristic algorithm is to expand the node with the highest incremental influence value, ℎ(𝑣), to maximize exposure for two campaigns.

To achieve a better computational accuracy, the algorithm applies Monte Carlo simulation to approximate each node’s influence impact by averaging results from multiple simulated influence spreads. Furthermore, to improve efficiency, the algorithm shall avoid recalculating the entire influence spread from scratch. It should only compute the changes in active and exposed nodes for each candidate node. This is achieved by diffusing influence to only a depth deeper and incrementally updates active and exposed sets.

In this proposed influence diffusion process, each incremental computation is conducted to a depth of one, focusing solely on evaluating the immediate neighbors of the current source node. This approach does not lead to issues with activation or exposure, as the algorithm will subsequently iterate through all nodes, prompting each node to activate its own neighbors. Consequently, there is no need to explore deeper from the initial node, as doing so would introduce unnecessary redundancy in the checks performed.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Algorithm 2: Incremental Influence Diffusion** | | | | | | | |
|  | **Input**: *graph*, *src* (source node), *active*, *exposed*, *campaign* | | | | | | |
|  | **Output**: *active\_increment* (set of newly activated nodes), *exposed\_increment* (set of newly exposed nodes) | | | | | | |
| 1 | Initialize *active\_increment* as an empty set | | | | | | |
| 2 | Initialize *exposed\_increment* as an empty set | | | | | | |
| 3 | **if** not *active*.contains(*src*) **then** | | | | | | |
| 4 |  | *active\_increment*.Add(*src*) | | | | | |
| 5 | **if** not *exposed*.contains(*src*) **then** | | | | | |
| 6 |  | *exposed\_increment*.Add(*src*) | | | | | |
| 7 | **for** *neighbo*r, *probab* in *graph*.GetAdjList(*src*) **do** | | | | | | |
| 8 |  | **if** not *exposed*.Contains(*neighbor*) **then** | | | | | |
| 9 |  |  | *exposed\_increment*.Add(*neighbor*) | | | | |
| 10 |  | **if** not *active*.Contains(*neighbor*) **then** | | | |
| 11 |  |  | **if** *probab* ≥ RandomValue() **then** | |
| 12 |  |  |  | *active\_increment*.Add(*neighbor*) | | | |
| 13 | **return** *active\_increment*, *exposed\_increment* | | | | | | |

By limiting the depth to one at this stage, we streamline the process and enhance computational efficiency while ensuring comprehensive coverage of the network in subsequent iterations. Nonetheless, the aforementioned influence diffusion with exhaustive BFS is still utilized in our final algorithm, but only once for each campaign.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Algorithm 3: IEMP Monte Carlo Greedy Heuristic** | | | | | | |
|  | **Input**: *graph*, *initial1* (*I1*), *initial2* (*I2*), *budget* (*k*), *rep* (number of simulations) | | | | | |
|  | **Output**: *balanced1* (*S1*), *balanced2* (*S2*) | | | | | |
| 1 | *num\_nodes* ← *graph*.GetNumNodes() | | | | | |
| 2 | Initialize *balanced1* as an empty set | | | | | |
| 3 | Initialize *balanced2* as an empty set | | | | | |
| 4 | **while** *balanced1*.Length() *+ balanced2*.Length() < *budget* **do** | | | | | |
| 5 |  | Initialize *h1\_rec* as an empty array | | | | |
| 6 |  | Initialize *h2\_rec* as an empty array | | | | |
| 7 |  | **for** *j* ← 0 to *rep*-1 **do** | | | | |
| 8 |  |  | *union1* ← *initial1* ∪ *balanced1* | | | |
| 9 |  |  | *union2* ← *initial2* ∪ *balanced2* | | | |
| 10 |  |  | *active1*, *exposed1* ← InfluenceDiffusionViaBFS(*graph*, *union1*, 1) | | | |
| 11 |  |  | *active2*, *exposed2* ← InfluenceDiffusionViaBFS(*graph*, *union2*, 2) | | | |
| 12 |  |  | *phi\_s1\_s2* ← ComputePhi(*num\_nodes*, *exposed1*, *exposed2*) | | | |
| 13 |  |  | **for** *i* ← 0 to *num\_nodes*-1 **do** | | | |
| 14 |  |  |  | **if** not *balanced1*.contains(*i*) and not *balanced2*.contains(*i*) **then** | | |
| 15 |  |  |  |  | *active1\_increment*, *exposed1\_increment* ← IncrementalInfluenceDiffusion( *graph*, *i,* *active1*, *exposed1*, 1) | |
| 16 |  |  |  |  | *active2\_increment*, *exposed2\_increment* ← IncrementalInfluenceDiffusion( *graph*, *i,* *active2*, *exposed2*, 2) | |
| 17 |  |  |  |  | *phi\_s1vi\_s2* ← ComputePhi(*num\_nodes*, *exposed1* ∪ *exposed1\_increment*, *exposed2*) | |
| 18 |  |  |  |  | *phi\_s1\_s2vi* ← ComputePhi(*num\_nodes*, *exposed*1, *exposed2* ∪ *exposed2\_increment*) | |
| 19 |  |  |  |  | *h1\_rec*[*i*] ← *h1\_rec*[*i*] + *phi\_s1vi\_s2 - phi\_s1\_s2* | |
| 20 |  |  |  |  | *h2\_rec*[*i*] ← *h2\_rec*[*i*] + *phi\_s1\_s2vi - phi\_s1\_s2* | |
| 21 |  | **for** *j* ← 0 to *num\_*nodes-1 **do** | | | |
| 22 |  |  | *h1\_rec*[*j*] ← *h1\_rec*[*j*] / *rep* | | |
| 23 |  |  | *h2\_rec*[*j*] ← *h2\_rec*[*j*] / *rep* | | |
| 24 |  | *new\_v1* ← IndexOfMaxElement(*h1\_rec*) | | | |
| 25 |  | *new\_v2* ← IndexOfMaxElement(*h2\_rec*) | | | |
| 26 |  | **if** *new\_v1* ≥ *new\_v2***then** | | | |
| 27 |  |  | *balanced1*.Add(*new\_v1*) | | |
| 28 |  | **else** | | | |
| 29 |  |  | *balanced2*.Add(*new\_v2*) | | |
| 30 | **return** *balanced1*, *balanced2* | | | | | |

## Evolutionary Approach

In this section, we address the IEMP by employing a genetic algorithm (GA) to identify optimal intervention sets *S1*​ and *S2* that maximize the expected influence spread within a network. Given the combinatorial complexity and large search space of this problem, optimization techniques like genetic algorithms provide an effective solution method. GAs are well-suited for this task due to their capability to explore complex, multimodal search spaces and to find near-optimal solutions within reasonable computational time. The GA approach is structured as follows:

* **Solution Representation**: Each candidate solution is represented by a binary vector, 𝑥 = {𝑥*1*, 𝑥*2*, …, 𝑥*|V|* , 𝑥*|V+1|*, 𝑥*|V+2|*, …, 𝑥*|V+V|*} where 𝑥*i* ∈ {0, 1}.
* **Fitness Function**: To guide the search towards optimal solutions, a fitness function is defined that evaluates each candidate solution based on the influence spread it achieves. The fitness function distinguishes between feasible and infeasible solutions.
* **Genetic Algorithm Operations**: In our approach, diversity is cherished; hence, the GA operations below is full of randomness, which makes this study quite experimental.

1. **Initialization**: The initial population is generated in three distinct groups to balance diversity and quality: The random population is created with random binary vectors where nodes are randomly added to intervention sets *S1* or *S2.* The controlled random population has nodes selected within a controlled range, increasing the probability of feasible solutions within budget. The ideal population is generated with exactly the budgeted number of nodes, creating solutions that strictly satisfy constraints.
2. **Survivor Selection and Offspring Generation**: Survivor selection is relatively simple. It involves choosing the top-performing individuals and a few mid-range and lower-performing solutions to maintain genetic diversity. Choosing pairs for breeding is done through random pair selection, except some controlled pairs for breeding among the best parents. Offspring generation is achieved through two-point crossover and mutation, where selected individuals exchange segments of their binary vectors to produce new solutions. Flip-bit mutation is applied probabilistically to encourage diversity, with a defined mutation rate applied to randomly chosen bits in the binary vectors of offspring solutions. This technique introduces further variation into the population and prevents premature convergence.

|  |  |  |  |
| --- | --- | --- | --- |
| **Algorithm 4: IEMP Genetic Algorithm** | | | |
|  | **Input**: *graph*, *initial1*, *initial2*, *budget* | | |
|  | **Output**: *balanced1*, *balanced2* | | |
| 1 | *HALT* ← 0.98 | | |
| 2 | *num\_nodes* ← *graph*.GetNumNodes() | | |
| 3 | *gen0* ← InitGen0(graph, *budget*) | | |
| 4 | Evaluate fitness of *gen0* | | |
| 5 | Sort *gen0* based on fitness value | | |
| 6 | *current\_best\_solution, current\_best\_solution\_val* ← RetrieveBestSolution(*gen0*) | | |
| 7 | *current\_gen* ← *gen0* | | |
| 8 | **for** *generation* ← 0to *gen*-2 **do** | | |
| 9 |  | **if** *current\_best\_solution\_val* / *num\_nodes* ≥ HALT**then** | |
| 10 |  |  | **break** |
| 11 |  | *parent\_gen* ← ChooseSurvivor(*current\_gen*) | |
| 12 |  | *next\_gen* ← GenerateOffSpring(*parent\_gen*) | |
| 13 |  | Evaluate fitness of *next\_gen* | |
| 14 |  | Sort *next\_gen* based on fitness value | |
| 15 |  | *candidate\_best\_solution, candidate\_best\_solution\_val* ← RetrieveBestSolution(*next\_gen*) | |
| 16 |  | **if** *candidate\_best\_solution\_val* ≥ *current\_best\_solution\_val* **then** | |
| 17 |  |  | *best\_solution* ← *candidate\_best\_solution* |
| 18 |  |  | *best\_solution\_val* ← *candidate\_best\_solution\_val* |
| 19 | *balanced1, balanced2* ← ConvertBinaryRepresentationToSetRepresentation( *best\_solution*) | | |
| 20 | **return** *balanced1*, *balanced2* | | |

Note that we may terminate the algorithm as soon as the halt condition is satisfied, which is defined as checking if Φ(*S1*, *S1*) is 98% as big as the total number of nodes. This allows us to improve the computational time by accepting results that are sufficiently close to optimal.

# 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

The template will number citations consecutively within brackets [1]. The sentence punctuation follows the bracket [2]. Refer simply to the reference number, as in [3]—do not use “Ref. [3]” or “reference [3]” except at the beginning of a sentence: “Reference [3] was the first ...”

Number footnotes separately in superscripts. Place the actual footnote at the bottom of the column in which it was cited. Do not put footnotes in the abstract or reference list. Use letters for table footnotes.

Unless there are six authors or more give all authors’ names; do not use “et al.”. Papers that have not been published, even if they have been submitted for publication, should be cited as “unpublished” [4]. Papers that have been accepted for publication should be cited as “in press” [5]. Capitalize only the first word in a paper title, except for proper nouns and element symbols.

For papers published in translation journals, please give the English citation first, followed by the original foreign-language citation [6].

1. G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955. *(references)*
2. J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
3. I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.
4. K. Elissa, “Title of paper if known,” unpublished.
5. R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.
6. Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].
7. M. Young, The Technical Writer’s Handbook. Mill Valley, CA: University Science, 1989.

**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.**