GUDIE

A flexible, user-defined method to extract subgraphs of interest from large graphs

Maria Inês Silva, David Aparício, Beatriz Malveiro Jorge, João Tiago Ascensão, Pedro Bizarro

{ maria.silva, david.aparicio, beatriz.malveiro, joao.ascensao, pedro.bizarro } @ feedzai.com

graph expansions | fraud detection | message passing algorithms | banking networks | anti-money laundering

feedzai

Motivation

Large, dense, small-world networks often emerge from social phenomena, including financial networks, social media, or epidemiology. As networks grow in importance, it is often necessary to partition them into meaningful units of analysis. However, what is considered meaningful for analysis is data- and use-case specific.
GUDIE uses user-defined criteria to retrieve the relevant context in order to accelerate and extend analysis capabilities in finance and other critical networks.

GUDIE

GUDIE is a message-passing algorithm that returns the most interesting expansion for each node in a list of nodes, based on user-defined interest.

High-Interest Nodes:

- Nodes that are closer to the seed
- Nodes in high-interest areas
- Nodes connecting the seed to high-interest areas
- Node interest does not necessarily increase with the node degree

User-Defined Interest

GUDIE considers user-defined criteria for the expansions to ensure flexibility and adaptability to different networks. GUDIE's expansions rely upon user-defined interest functions and the structural properties of the graph.

VUDIE: node or **v**ertex interest $V(G) \rightarrow [0, 1]$

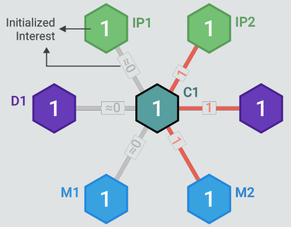
LUDIE: edge or **l**ink interest $E(G) \rightarrow [0, 1]$

Graph G
Seed Nodes → **GUDIE** → **GraphUnits for each seed**
(list of interesting subgraphs)

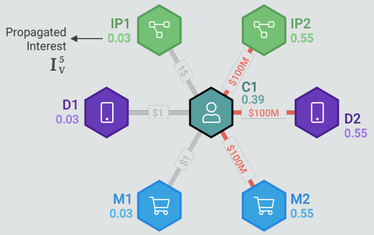
VUDIE, LUDIE, Number of hops
Interest propagation function, Interest aggregation function,
Interest expansion tolerance, Distance decay function

Method

1. **Initialization**
   Node and edge interest scores, I_V and I_E, respectively, are computed according to the interest functions VUDIE, V(G), and LUDIE, E(G). At this point, every node and edge in the graph G has an interest score.

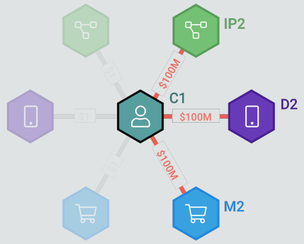


2. **Interest Propagation**
   Nodes message their neighbors their interest score. The interest propagation process runs for **h** hops. According to the **interest propagation function**, nodes split their node interest among their neighbors, and update their interest according to the **interest aggregation function**. After this step, we have a new propagated interest score I_V^h .



3. **Seed Expansion**
   The seed expansion process runs on top of G and the propagated node interest I_V^h . The expansion starts by computing the minimum allowed interest for the seed using the **interest expansion tolerance**. During expansion, distant nodes from the seed are penalized according to the **distance decay function**. At the end of the seed expansion process, each node contains the expansions traversing them.

4. **Obtain GraphUnits**
   GUDIE uses a map-reduce operator to obtain the subgraphs.



5. **Network Analysis**
   GraphUnits can assist complex linked analysis by enabling more relevant and concise expansions.

ECML PKDD 2021
VIRTUAL
13-17 September