# Large-scale multi-objective influence maximisation with network downscaling

Elia Cunegatti [1,2] Giovanni Iacca [1] Doina Bucur [2] [1]University of Trento - [2]University of Twente

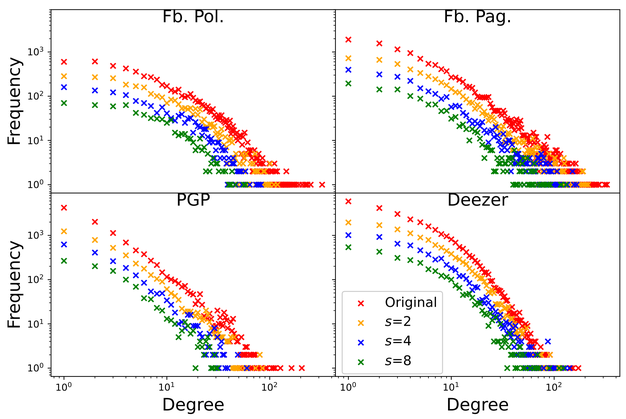
## Introduction

We devise a **graph summary** method to tackle the **Influence Maximisation (IM)** problem on *social networks*. Starting from a graph $G = (V, E)$, where $V$ is the set of nodes and $E$ is the set of edges in $G$, IM is the problem of finding a seed set $S$ s.t $|S| \leq k$ that maximise the *influence propagation* over $G$, given a certain propagation model. The used propagation models are:

- Independent Cascade (IC): probability *p* is equal across all links, and is fixed;
- Weighted Cascade (WC): *p* is inversely proportional to the number of links for each node $m \in V$, i.e., $p = \frac{1}{deg(m)}$

## Downscaling

- preserves the number of **communities** in the network;
- downscales **the number of nodes and edges** by a *scale factor* $s$;
- preserves the node **degree distribution**.

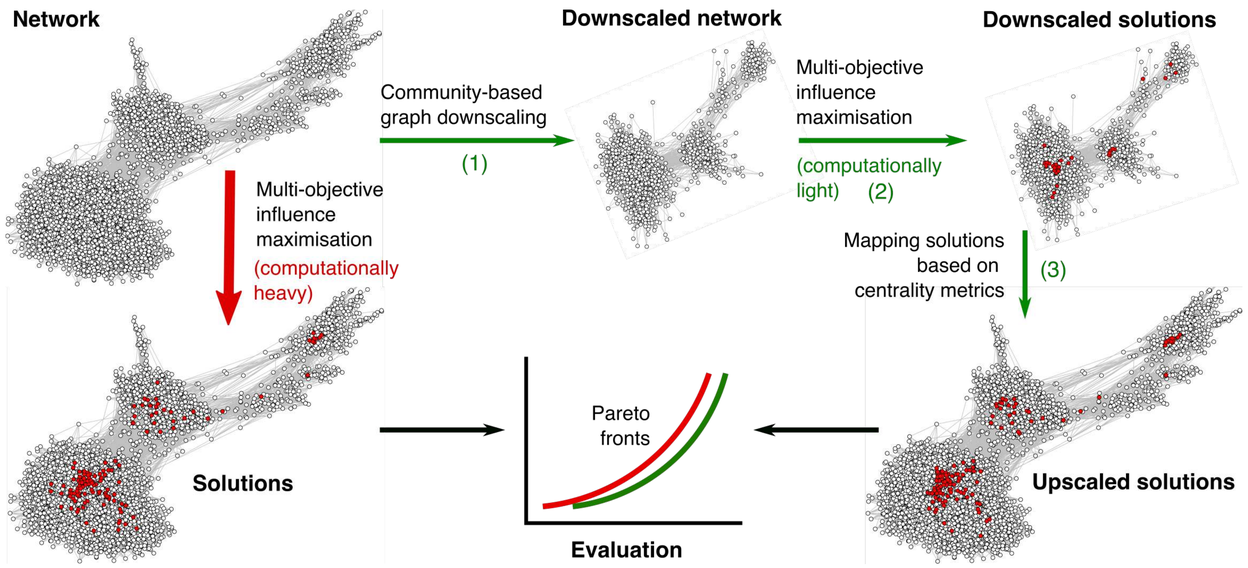



Social networks follow *power-law distribution* in which the fraction of nodes with a certain degree $d$ is proportional $d^{-\alpha}$ where $\alpha > 0$.

## Method

The proposed method is divided into three main steps:

1. **Downscaling**;
2. **Multi-objective Evolutionary Algorithm (MOEA)**;
3. **Upscaling**.




## Multi-objective Evolutionary Algorithm (MOEA)

We use **NSGA-II** with tournament selection, one-point crossover and random mutation with the boost of *smart-initialization*.

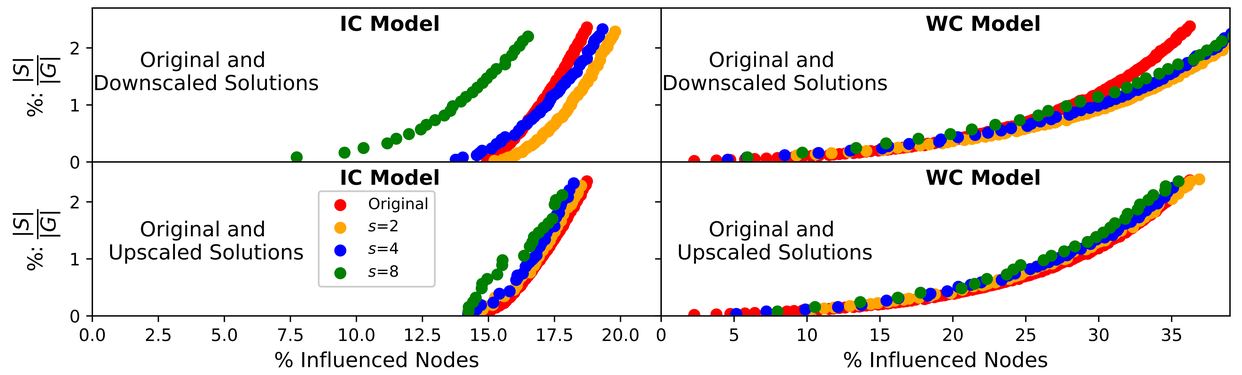
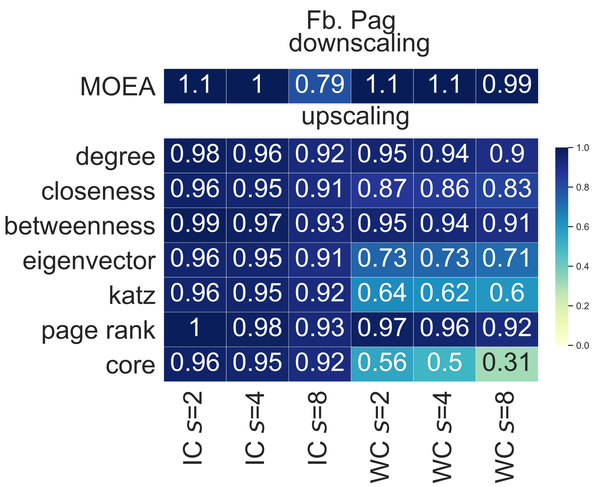



Figure 1. Original and Upscaled (PageRank) PFs for Fb. Pag.

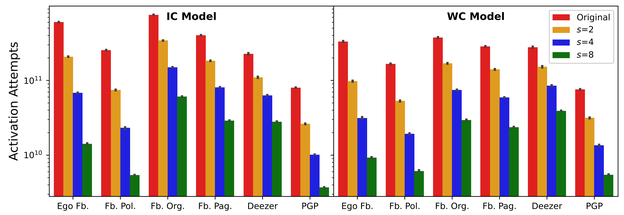
## Upscaling

This step takes in input two graphs $G$ (unscaled) and $G_s$ (downscaled) and a set seed on $G_s$, denoted as $S_s$. The task is to translate $S_s$ into a seed set $S$ on $G$. We find a **matching seed set** $S$ of $|S_s| \times s$ nodes in $G$. We do this per community. Each node in $S_s$ has a **rank in its community**, based on the **centrality** values of all nodes in that community.




## Runtime

**Runtime decreases by a factor from 2x to 5x** when the $s$ is doubled.




## Conclusion

- $\alpha$ preserved in the downscaling;
- **PageRank** best centrality for upscaling;
- **Hypearea** in $[0.78, 1]$ for $s \in [2, 4, 8]$;
- **Diversity** in the solutions due to the MOEA;
- Runtime massively decreased.

Paper: Code:



elia.cunegatti@unitn.it