Influence Mining Framework

# Problem Statement

Budget alone is not the only parameter, time, influence and other factors may be involved. The influence and resilience benchmarking methods run in seclusion.

# Objectives

1. Identify which method performs best in influence on minimum budget.
2. Identify which method performs best in influence on maximum budget.
3. Identify which method performs best in influence on minimum budget.
4. Identify which method is most efficient on minimum budget.
5. Identify which method is most efficient on maximum budget.
6. Identify which method performs best in influence early on time.
7. Identify which method performs best in influence late on time.
8. Identify which method reaches desired influence first on low budget.
9. Identify which method reaches desired influence last on low budget.
10. Identify which method reaches desired influence first on high budget.
11. Identify which method reaches desired influence last on high budget.
12. Are the state-of-art methods best performers in all segments?
13. Are the results from state-of-art methods consistent on all networks?

# Study Design

## Data set

#### Network statistics

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **graph** | **edges** | **nodes** | **ratio** | **avg\_deg** | **apl** | **high\_dev** | **diameter** | **transitivity** | **assortativity** | **avg\_dist** | **triads** | **power\_law** |
| karate | 78 | 34 | 0.436 | 4.59 | 2.408 | 17 | 5 | 0.25568 | -0.476 | 2.408 | 135 | 2.13 |
| football | 616 | 115 | 0.187 | 10.71 | 2.508 | 13 | 4 | 0.40724 | 0.142 | 2.508 | 2430 | 21.19 |
| world\_trade | 1000 | 80 | 0.080 | 25.00 | 1.724 | 95 | 3 | 0.45957 | -0.294 | 1.724 | 13299 | 2.56 |
| nematode | 2359 | 297 | 0.126 | 15.89 | 2.455 | 139 | 5 | 0.18071 | -0.150 | 2.455 | 9723 | 3.29 |
| automata | 2770 | 1096 | 0.396 | 5.05 | 1.995 | 1096 | 2 | 0.00556 | -0.746 | 1.995 | 6987 | 3.46 |
| author | 914 | 379 | 0.415 | 4.82 | 6.042 | 34 | 17 | 0.43058 | -0.082 | 6.042 | 2763 | 3.36 |
| protein | 3941 | 1458 | 0.370 | 5.41 | 6.812 | 112 | 19 | 0.05178 | -0.207 | 6.812 | 618 | 3.24 |
| influence\_citation | 1993 | 1254 | 0.629 | 3.18 | 3.806 | 172 | 6 | 0.05710 | -0.407 | 3.806 | 3537 | 2.35 |
| routers | 6632 | 2113 | 0.319 | 6.28 | 4.607 | 109 | 12 | 0.22826 | 0.019 | 4.607 | 31212 | 2.74 |
| hamster | 12476 | 1788 | 0.143 | 13.96 | 3.453 | 272 | 14 | 0.09037 | -0.089 | 3.453 | 50223 | 4.32 |
| moreno | 12969 | 2539 | 0.196 | 10.22 | 4.559 | 36 | 10 | 0.14189 | 0.288 | 4.559 | 14082 | 8.42 |
| fb\_tvshows | 17262 | 3892 | 0.225 | 8.87 | 6.276 | 126 | 20 | 0.59064 | 0.561 | 6.276 | 261270 | 2.75 |
| political | 19089 | 1222 | 0.064 | 31.24 | 2.738 | 468 | 8 | 0.22596 | -0.196 | 2.738 | 303129 | 3.89 |
| usairport | 28235 | 1572 | 0.056 | 35.92 | 3.115 | 596 | 8 | 0.38414 | -0.122 | 3.115 | 735516 | 1.86 |
| spam | 37375 | 4767 | 0.128 | 15.68 | 3.794 | 477 | 12 | 0.14519 | 0.000 | 3.794 | 387051 | 2.42 |
| arxiv | 41602 | 6794 | 0.163 | 12.25 | 5.780 | 341 | 16 | 0.19831 | 0.376 | 5.780 | 45000 | 4.20 |
| ita2000 | 67195 | 3304 | 0.049 | 40.67 | 4.037 | 1826 | 13 | 0.24894 | 0.007 | 4.037 | 302496 | 1.69 |
| jdk6 | 150985 | 6434 | 0.043 | 46.93 | 2.120 | 32530 | 7 | 0.01110 | -0.223 | 2.120 | 584526 | 2.01 |
| johnshopkins | 186572 | 5157 | 0.028 | 72.36 | 2.710 | 886 | 8 | 0.19316 | 0.080 | 2.710 | 4943352 | 5.57 |
| city2010 | 39904 | 1209 | 0.030 | 66.01 | 2.287 | 1045 | 5 | 0.34313 | -0.241 | 2.287 | 2021343 | 2.10 |
| city2013 | 49926 | 1358 | 0.027 | 73.53 | 2.268 | 1154 | 4 | 0.34883 | -0.251 | 2.268 | 2931834 | 6.12 |
| city2016 | 62523 | 1405 | 0.022 | 89.00 | 2.264 | 1188 | 5 | 0.38030 | -0.222 | 2.264 | 4223793 | 2.21 |
| city2019 | 68668 | 1422 | 0.021 | 96.58 | 2.244 | 1258 | 5 | 0.39310 | -0.206 | 2.244 | 4932006 | 5.93 |
| caida | 53381 | 26475 | 0.496 | 4.03 | 3.876 | 2628 | 17 | 0.00732 | -0.195 | 3.876 | 109095 | 2.09 |
| linux | 230201 | 30817 | 0.134 | 14.94 | 3.254 | 9775 | 12 | 0.00282 | -0.175 | 3.254 | 512586 | 1.96 |
| youtube | 293360 | 94238 | 0.321 | 6.23 | 4.361 | 7593 | 11 | 0.00105 | -0.060 | 4.361 | 75423 | 2.45 |
| epinions | 508836 | 75877 | 0.149 | 13.41 | 4.308 | 3079 | 15 | 0.06568 | -0.011 | 4.308 | 4873443 | 1.70 |
| wordnet | 656230 | 145145 | 0.221 | 9.04 | 5.532 | 1008 | 16 | 0.09575 | -0.063 | 5.532 | 3433944 | 2.47 |

#### Key notes

1. Power law coefficient does not stay consistent. In 2010 and 2016, this value is comparable with social networks, but not for 2013 and 2019. It is important to note that using power-law to study scale free structure of networks has been questioned by researchers (cite *The Anatomy of the Facebook Social Graph*, Johan Ugander AND *Power-law distributions in empirical data*, Aaron Clauset). The power-law plots illustrate fat tail, proving the scale-free structure of networks.

## Degree Distribution

|  |  |
| --- | --- |
| karate |  |
| football |  |
| world\_trade |  |
| nematode |  |
| automata |  |
| author |  |
| protein |  |
| influence\_citation |  |
| routers |  |
| hamster |  |
| moreno |  |
| fb\_tvshows |  |
| political |  |
| usairport |  |
| spam |  |
| arxiv |  |
| ita2000 |  |
| jdk6 |  |
| johnshopkins |  |
| city2010 |  |
| city2013 |  |
| city2016 |  |
| city2019 |  |
| caida |  |
| linux |  |
| youtube |  |
| epinions |  |
| wordnet |  |

## Power Law Fit

## Influence Mining Methods

## Influence Test Methods

1. Resilience test
2. Influence test under Linear Threshold model
3. Influence test under Independent Cascade model

## Experiments and Observations

### Accuracy of Methods

The influence of 1% most influential nodes extracted

The influence of 2% most influential nodes extracted

The influence of 3% most influential nodes extracted

The influence of 4% most influential nodes extracted

The influence of 5% most influential nodes extracted

### Efficiency of Methods

The influence of 1% most influential nodes extracted

The influence of 2% most influential nodes extracted

The influence of 3% most influential nodes extracted

The influence of 4% most influential nodes extracted

The influence of 5% most influential nodes extracted

**Observations:**

1. Degree is almost identical to Adaptive Degree