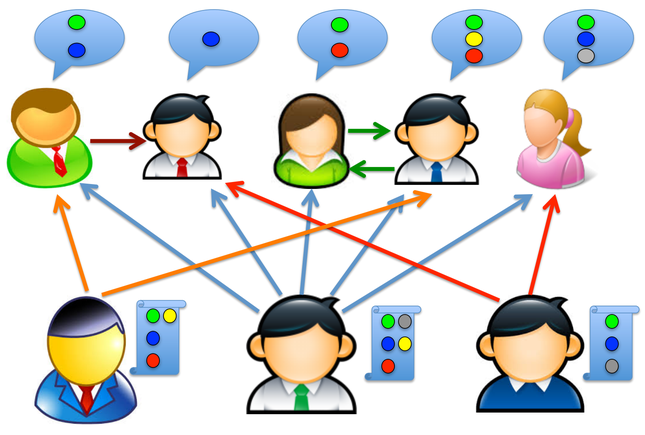
On the Efficiency of the Information Networks in Social Media

Mahmoudreza Babaei, Przemyslaw A. Grabowicz
Isabel Valera, Krishna P. Gummadi, Manuel Gomez-Rodriguez

MAX-PLANCK-GESELLSCHAFT

Max Planck Institute for Software Systems

1. Information Network



- Twitter
- Google Scholar
- Facebook
- Google +

2. Motivation

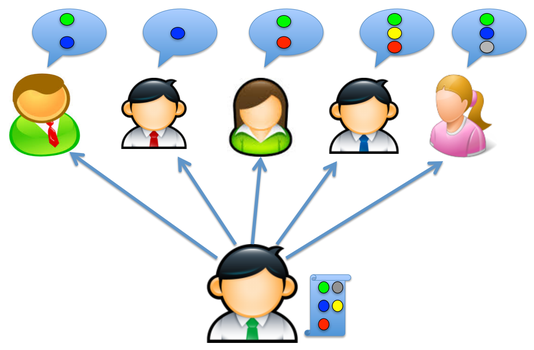
Social media users follow multiple other users to receive relevant information:

- 30% of Twitter users follow 50+ other users
- 50% of Twitter users receive 500+ tweets/day

Problems:

1. **Difficulties in managing large number of followees/subscriptions**
2. **Overload with redundant information**
3. **Delays in arrival of information**

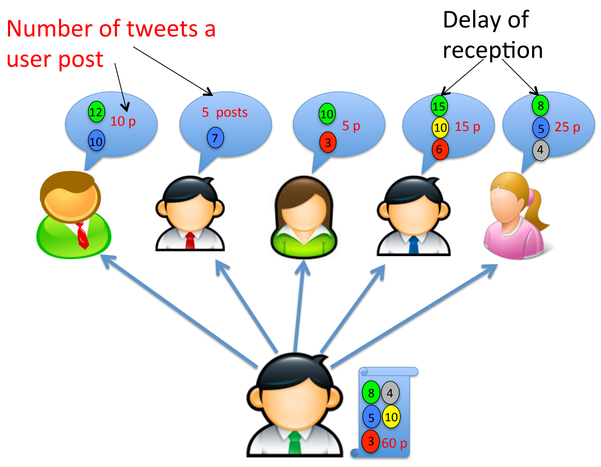
3. Key research question



- **How *efficient* are users at selecting their information sources?**
- **How we can improve the efficiency of users?**

4. Definitions of efficiency

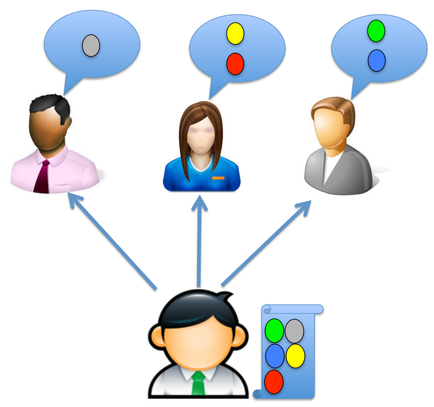
To compute an efficiency we compare the **original set** of followees with a **corresponding optimal set** that provides the **same pieces of information**.



The original set of followees

1. Link-optimal set

contains the smallest number of users.

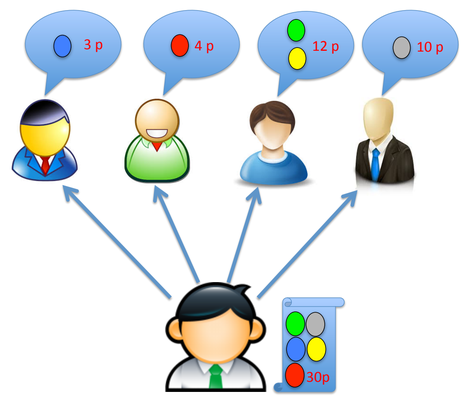


Minimum set contains 3 users, while original set contains 5 users:

Link efficiency $E_u^l = \frac{|U^l|}{|U|} = \frac{3}{5}$

2. Inflow-optimal set

provides the least amount of tweets per time unit.

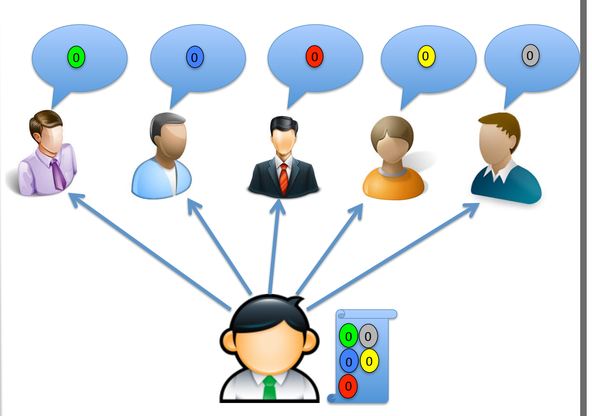


Optimal set contains 30 posts, while original set contains 60 posts:

Inflow efficiency $E_u^f = \frac{f(U^f)}{f(U)} = \frac{30}{60}$

3. Delay-optimal set

provides the contagions as early as possible.



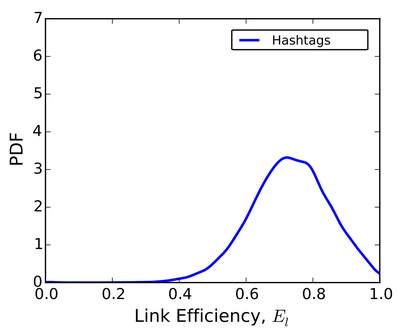
Optimal set contains users providing the contagions without delay:

Delay efficiency $E_u^t = \frac{1}{1+\langle t_i - t_i^0 \rangle_{i \in I_u}} = \frac{1}{1+30/5}$

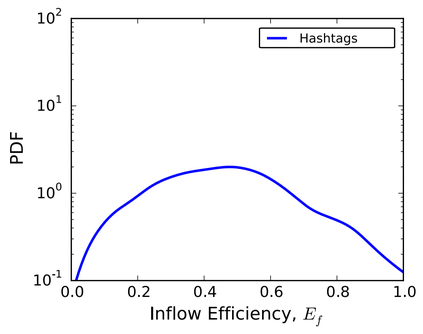
Efficiencies take **values from 0 to 1**, where **1 corresponds to a perfectly efficient user**.

5. Efficiency of users in real world

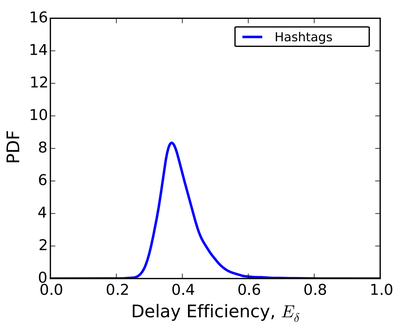
- **Users acquire information sub-optimally (efficiencies<1)**
- **Users tend to be less efficient at acquiring popular pieces of information**



Link Efficiency



Inflow Efficiency



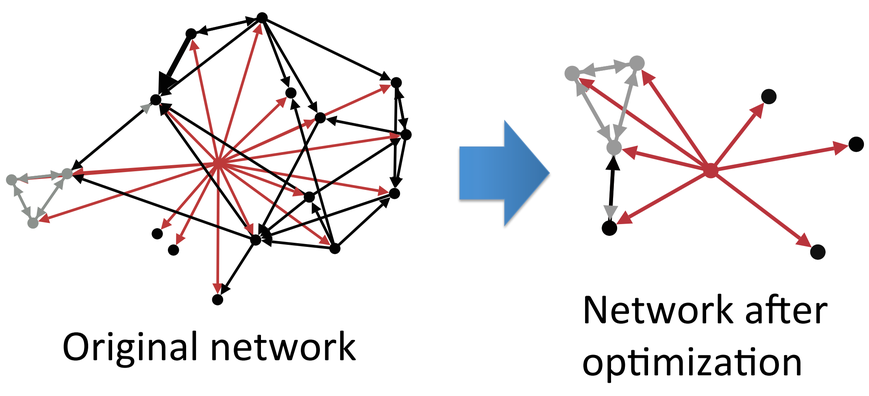
Delay Efficiency

We introduce a fast heuristic method that **cross-optimizes both inflow and delay efficiencies at once**, using a greedy algorithm solving weighted set cover problem.

Improved by 80%!

Improved by 40%!

6. Structure of optimized users' networks



Original network

Network after optimization

The optimized information networks **cannot be discovered via triangle closure** (creating links to friends of friends)

7. Future work

- **Could we creating a relink recommender system for real information networks that improves various efficiencies of users?**
- **Does user efficiency change over time? What is its relation with user activity?**
- **Compare user efficiencies across different existing information networks**