

Assessing User Engagement Capacity as a Driver of Reach of Online Health Platforms

Alexander Nikolaev Ph.D, SUNY at Buffalo ISE

Rahul Gopalsamy, SUNY at Buffalo ISE

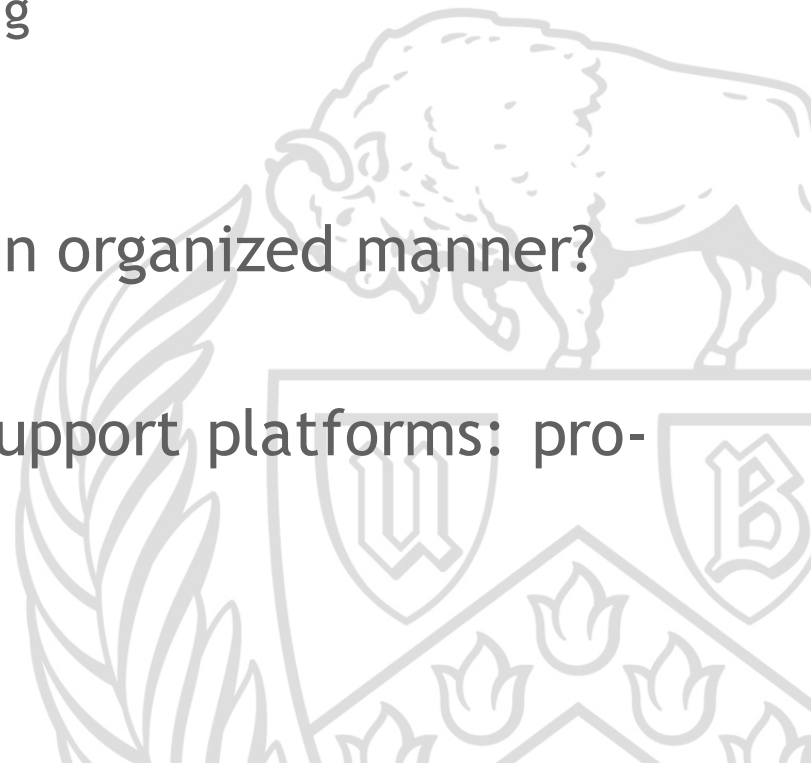
Alexander Semenov, University of Jyvaskyla

Eduardo Pasilliao, AFRL

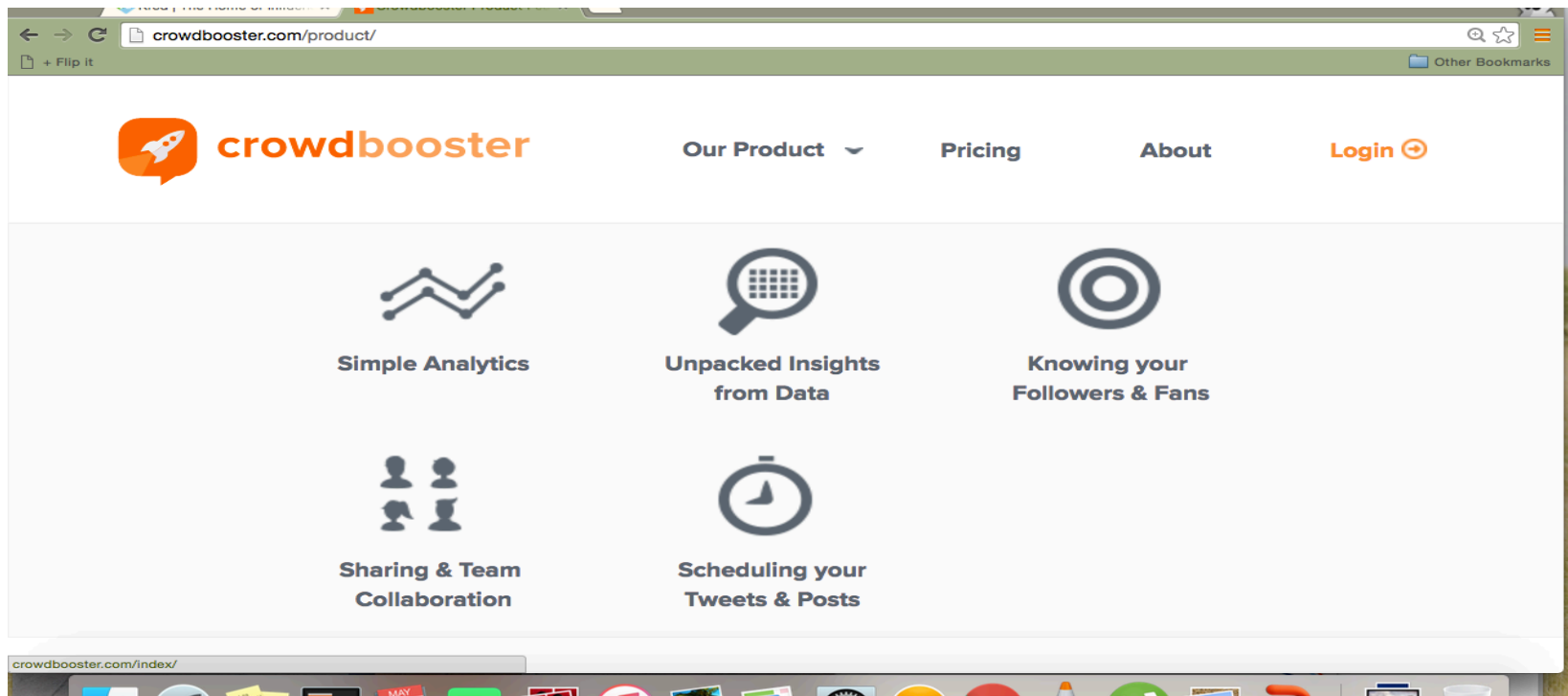


Online Community Building

- Consider a group that seeks to build its customer base or followship:
 - Products or services distribution
 - Interest or information sharing
 - Ideologies, beliefs adoption
- Can a group grow itself in an organized manner?
- The case of online social support platforms: pro-health, easy to observe



Example: Crowd-Boosting Services



Example: Engagement on Twitter



RE-AIM and Positive Externality Effects

On an online platform, positive network externalities occur in two instances:

- (1) when a new user joins the network and authors a new post, and
- (2) when an existing user contributes new user-generated content

Hypothesis: the internal growth of a platform (achieved through added user-generated content) leads to its external growth (the increase in the number of newly registered users): i.e., higher engagement leads to higher reach.

Outline

Engagement capacity measures the users' ability to engage their peers into contributing to the platform.

Targeted engagement capacity serves to assess how one group of users engages another group.

Engaging Team Formation Problem helps us find a group which maximizes the engagement.

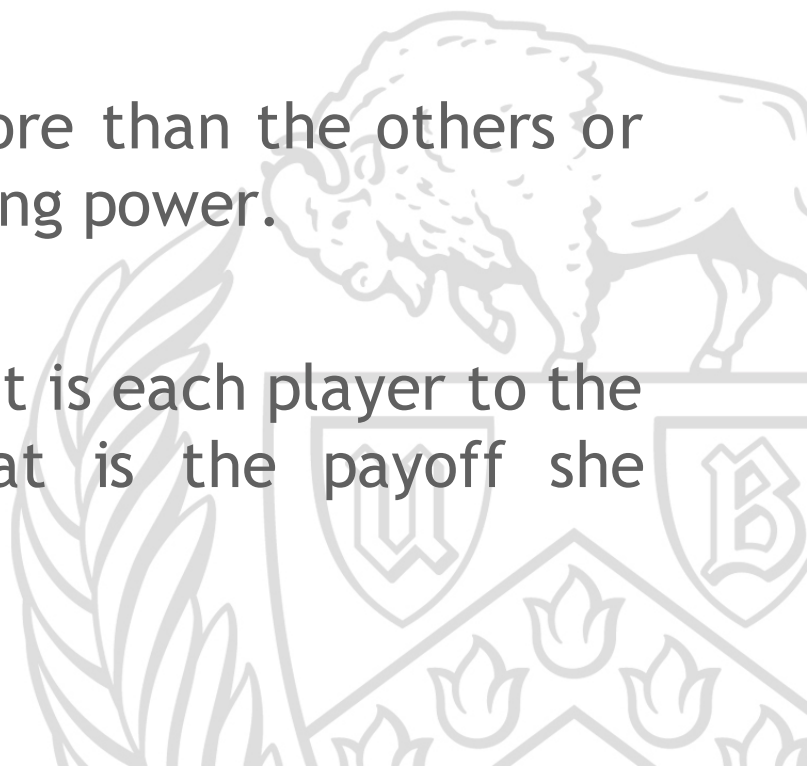
KDD Reviewers: do we have empirical evidence that engagement value correlates with reach metrics?

Cooperative Game Theory and Shapley Value

Suppose a set of players co-operate with each other and obtain a certain gain from that co-operation.

Some players may contribute more than the others or may possess different bargaining power.

In such a scenario, how important is each player to the overall cooperation and what is the payoff she should reasonably expect?



Shapley Value - Example

Consider 2 players A & B. If A produces 2 units when working alone, if B produces 1 unit while working alone and A & B together produce 5 units, what is the contribution of each in this case?

Ways in which collaborations can be initiated :

A -> A & B

B -> B & A

$S(A) = 3, S(B) = 2$

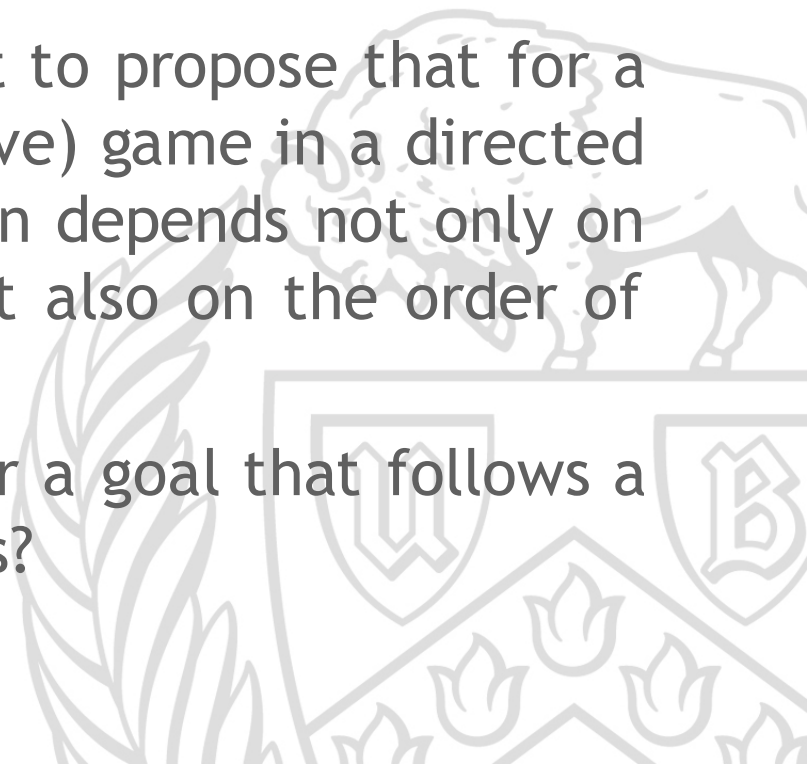


Graph based Cooperative Games

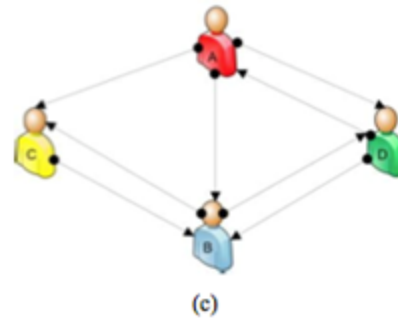
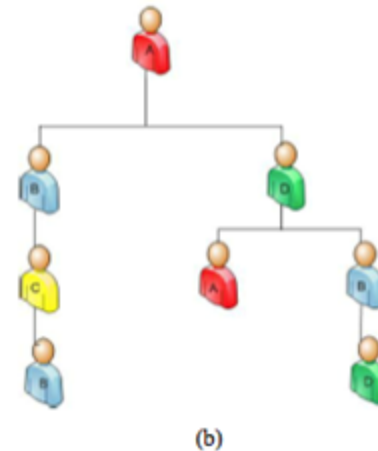
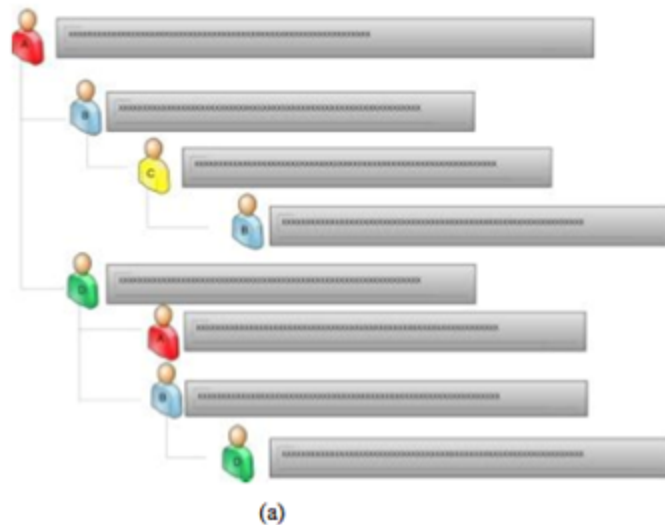
The first cooperative games on graphs were formulated by Myerson.

Nowak and Razdik were the first to propose that for a transferable utility (cooperative) game in a directed graph, the worth of a coalition depends not only on the coalition membership, but also on the order of members within the coalition.

- who is more responsible for a goal that follows a sequence of creative passes?



Forum communication networks

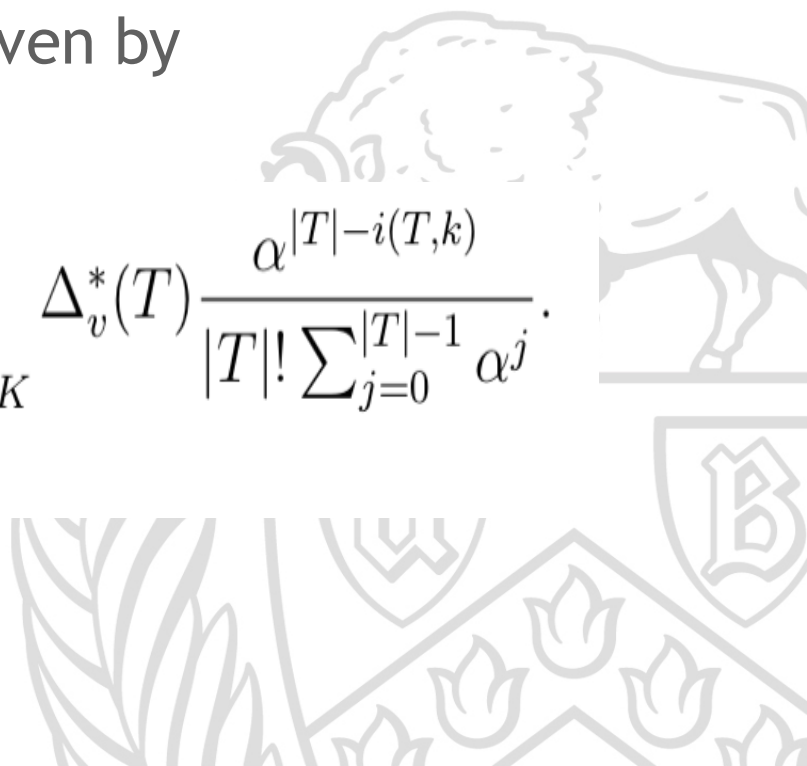


- Who started it?
- Who kept it going?

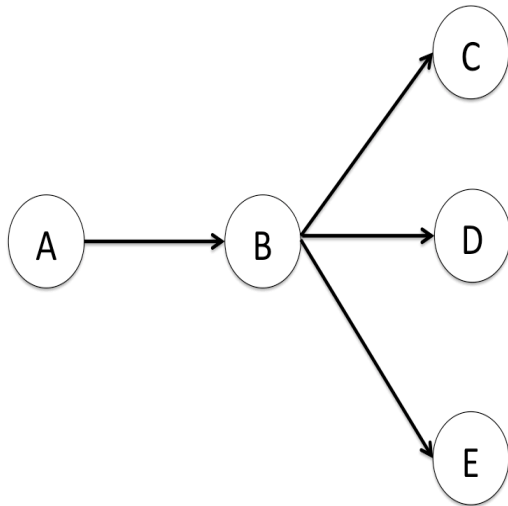
The Game of Engagement Generation

The value generated by a k-coalition is shared proportionally to the positions of the appearances of the coalition members and is given by

$$\Psi_i^{K-\alpha}(N, v) = \sum_{T \in \Omega^K(N), i \in H(T), k=1, \dots, K} \Delta_v^*(T) \frac{\alpha^{|T|-i(T,k)}}{|T|! \sum_{j=0}^{|T|-1} \alpha^j}.$$



An Example



- Forum – A collection of communication threads, contributed to by discussants interested in a particular topic.
- Thread – a sequence of communication
- Subthreads – A, AB, ABC, ABD, ABE
- Engaging subthreads – A, AB
- Engagement Capacity of User A is ...

Engagement Capacity Calculation Logic

Given a forum's snapshot (historical data), set K to be the largest number of posts contributed by the same user to any subthread.

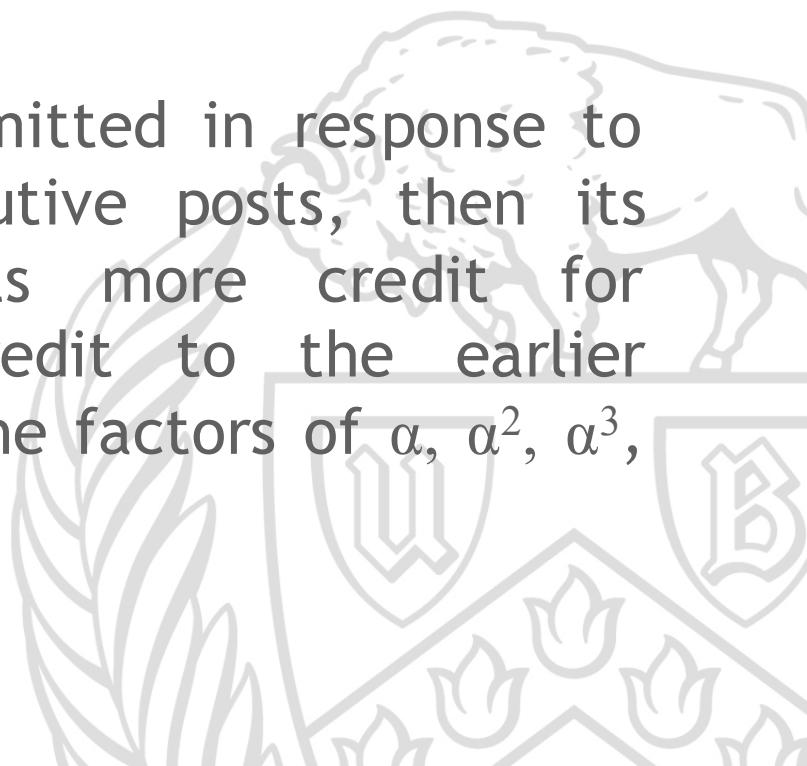
Set $\Delta_v^*(T)$ to return a total number of posts immediately succeeding such engaging subthreads $p \in P$ that have the same membership, size and structure as k -coalition $T \in \Omega^K(N)$.

The engagement capacity of forum user $i \in N$ is the value that returned for this user as a solution to the game (N, v, P) .

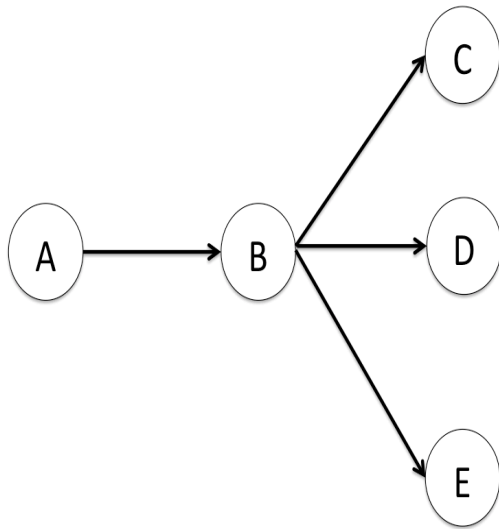
Engagement Capacity Calculation Logic

Note that the coefficient $\alpha \in [0, 1]$ in $\{\Psi^{K-\alpha}\}_{K \in I+, \alpha \in [0,1]}$ captures the engagement share tradeoff between thread contributors.

As such, if a new post is submitted in response to multiple (preceding) consecutive posts, then its immediate predecessor gets more credit for attracting it, with the credit to the earlier predecessors discounted by the factors of α , α^2 , α^3 , etc., respectively.



The Example



- Subthreads – A, AB, ABC, ABD, ABE
- Engaging subthreads – A, AB
- Engaging Capacity of User A

$$\eta_{A,(a)} = 1 + 3 * \frac{\alpha}{\alpha + 1}$$

- Engaging Capacity of User B

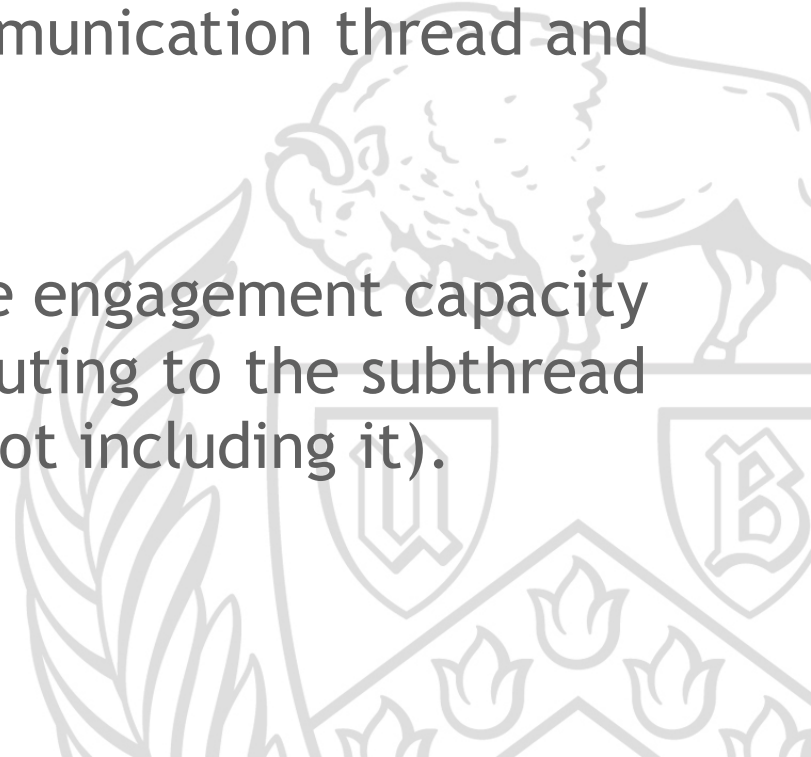
$$\eta_{B,(a)} = 3 * \frac{1}{\alpha + 1}$$

(If $\alpha = 1$, $\eta_{A,(a)}=2.5$ and $\eta_{B,(a)} = 1.5$)

Dynamic Updating

Each new user post submitted in response to another post, or sequence of posts, brings in one unit of engagement value to the communication thread and to the platform as a whole.

A newly added post increases the engagement capacity values of all the users contributing to the subthread leading to the new post (but not including it).

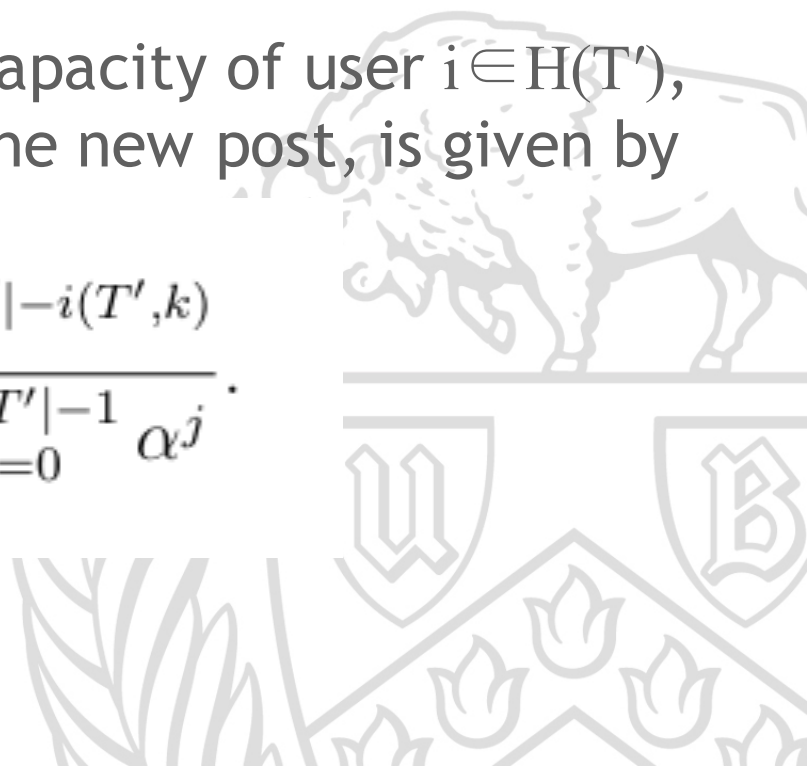


Dynamic Updating

Consider a k -coalition $T' \in \Omega^K(N)$ with the same membership, size and structure as this subthread.

The increase in the engagement capacity of user $i \in H(T')$, resulting from the addition of the new post, is given by

$$\Delta_i = \sum_{k=1,2,\dots,K} \frac{\alpha^{|T'| - i(T',k)}}{\sum_{j=0}^{|T'|-1} \alpha^j}.$$



Datasets and Data Collection

The forum contribution records were collected from an online healthcare platform.

It has about 200 social support forums and about as many “ask an expert” forums.

The website has close to 3 Million active and inactive threads and attracts about 8 Million visitors every month.

The users interact through discussion boards, contribute personal journal entries, and post notes on their friends home pages.

The data most relevant to the present study are those of the users’ interactions on discussion forums: such forums allow the users to give each other social

Datasets and Data Collection

A web crawler was used to collect the data of the contributed posts, and their sequences, in the “anxiety”, “cholesterol control” and “weight-loss & dieting” forums.

- The *anxiety forum* has about 33500 threads dating back to the early 2000s. The anxiety forum was chosen because it provides an example of a social support forum. People usually have a lot more conversations, especially one-to-one communication on such forums.
- The *cholesterol control forum* has about 300 threads. This forum is more focused and usually the answers are more to the point.
- The *weight-loss & dieting forum* has about 7000 threads. This is a mixed forum where people could give to the point answers or offer social support.

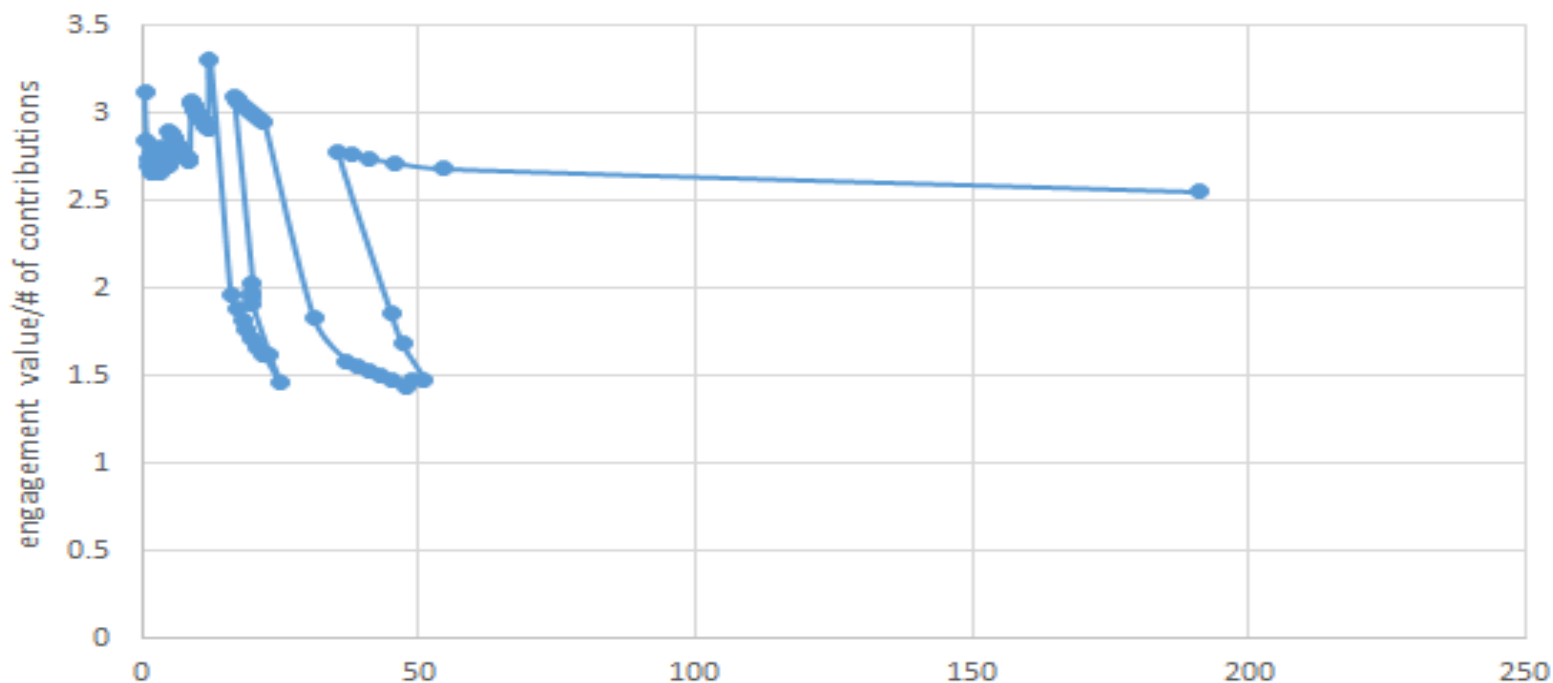
Datasets and Data Collection

Each thread consists of at least one post, however, not all the threads have replies; the unanswered threads do not affect the engagement capacity computation.

After cleaning up threads with no replies we end up with a database of about 48800 unique users.

A data tree is built for each analyzed thread. In each such tree, every user contribution is represented by a node; a directed link is drawn from node A to node B, if user B replies to user A.

Engagement Value per contribution on the Pro-Health Platform

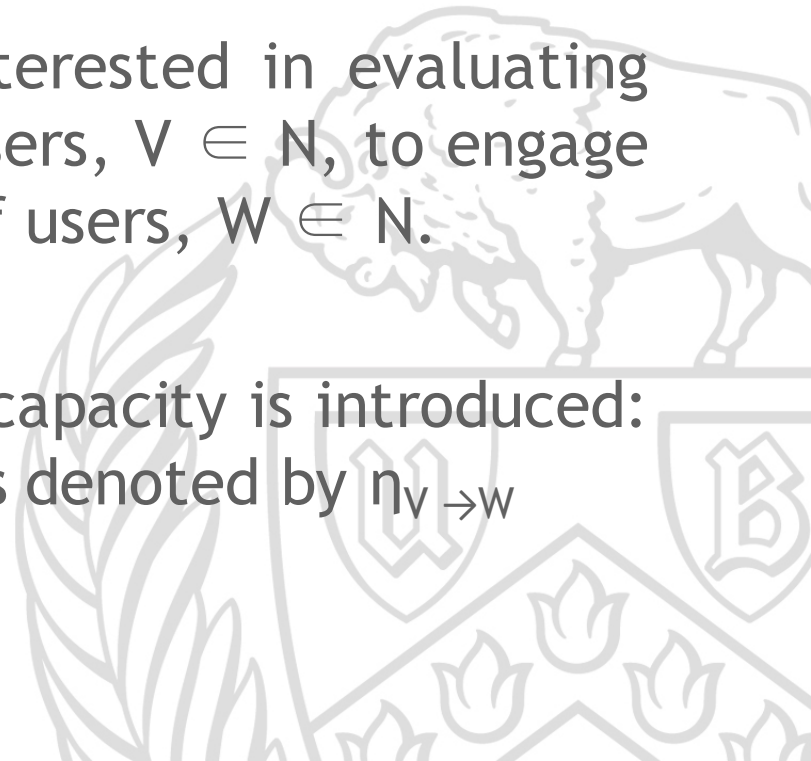


Targeted Engagement Capacity

How does one user manage to engage another fellow user?

More generally, one may be interested in evaluating the ability of a given set of users, $V \in \mathcal{N}$, to engage the members of another set of users, $W \in \mathcal{N}$.

The term targeted engagement capacity is introduced: then, the discussed quantity is denoted by $\eta_{V \rightarrow W}$



Targeted Engagement Capacity

Targeted Engagement capacity is defined as the sum of the shares allocated to the members of V in the game of engaging the members of W .

The targeted engagement capacity is given,

$$\eta_{V \rightarrow W} = \sum_{i \in V, j \in W} \sum_{i \in H(T), k=1, \dots, K} \Delta_v^*(T) \frac{\alpha^{|T| - i(T, k)}}{|T|! \sum_{j=0}^{|T|-1} \alpha^j},$$

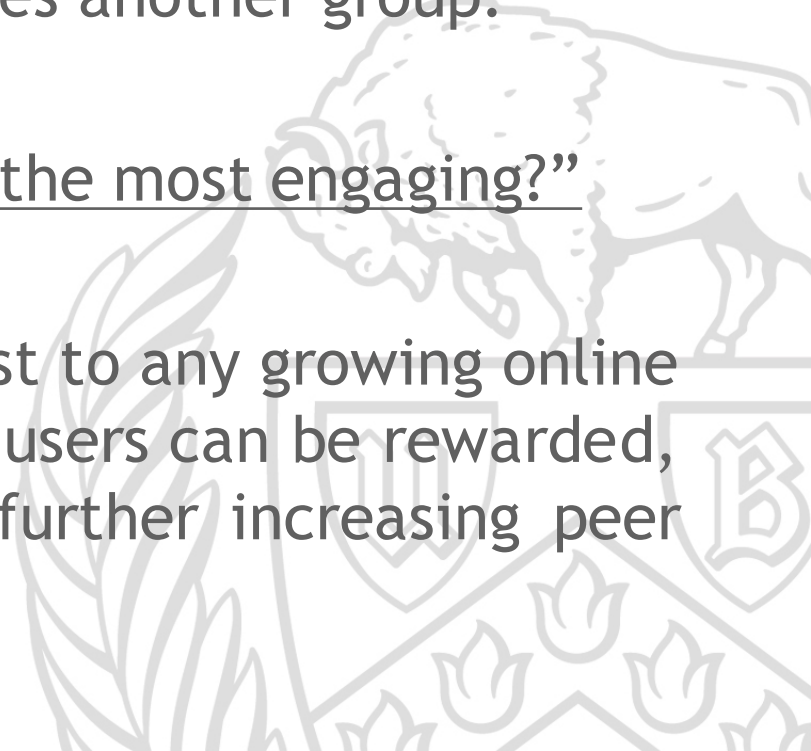
with the sum taken over the users in respective sets.

Engaging Team Formation Problem

The introduction of targeted engagement capacity allows one to formulate the problem of identifying a group of users that best engages another group.

“What group of users (team) is the most engaging?”

This question is of special interest to any growing online platform, since such teams of users can be rewarded, encouraged, and assisted in further increasing peer engagement and retention.

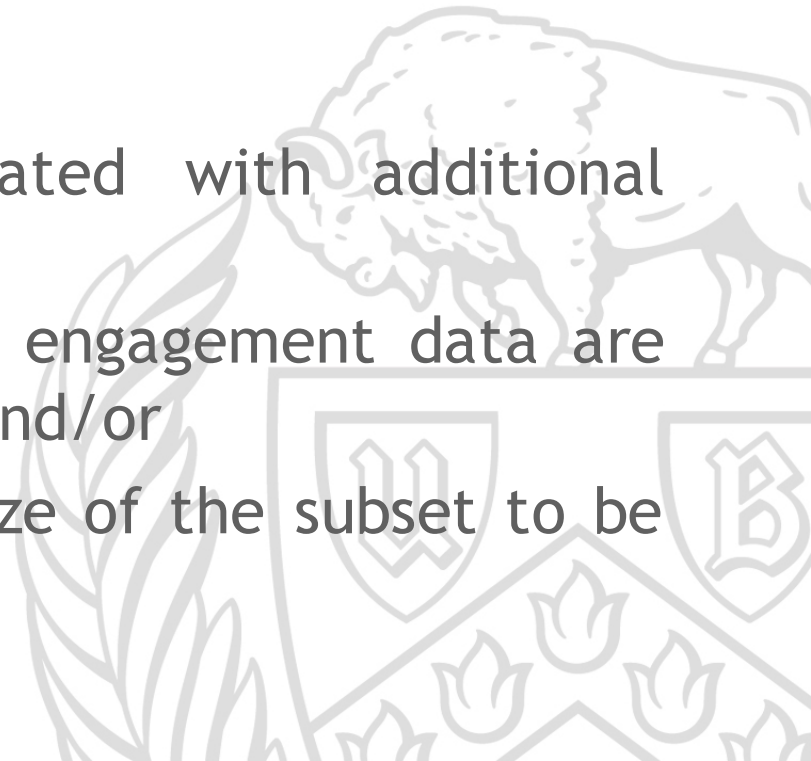


Engaging Team Formation Problem

The EngTFP objective is to select a subset of users with a maximal targeted engagement capacity towards all the other users.

The problem can be formulated with additional constraints, e.g.,

- reflecting which historical engagement data are to be taken into account, and/or
- reflecting the maximum size of the subset to be selected.



Engaging Team Formation Problem

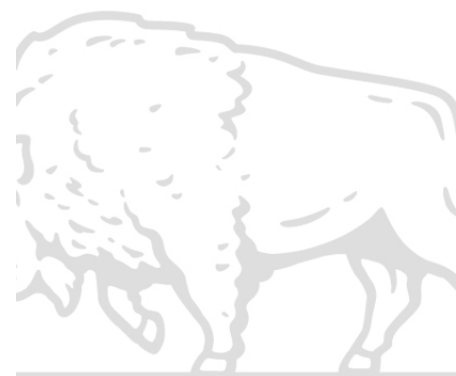
An MIP formulation

$$\max \sum_{i \in N} \sum_{j \in N, j \neq i} \eta_{i \rightarrow j} Y_{ij}$$

$$\text{s.t. } Y_{ij} \leq X_i, \quad \forall i, j,$$

$$Y_{ij} \leq 1 - X_j, \quad \forall i, j,$$

$$X_i, Y_{ij} \in \{0, 1\}, \quad \forall i, j.$$



What next?

Paved a way to conducting quantitative engagement analyses.

Developed techniques for computing user engagement capacity and targeted engagement capacity.

But does the hypothesis

“higher engagement leads to higher reach” hold?



Analysis of New User Arrivals

Analyze temporal, moving window-based, engagement capacity dynamics (e.g., per week); empirically validate the claim that higher engagement leads to higher reach

The objective is to gauge the correlation between the engagement value dynamics and:

- addition of new users
- delurking



Approach

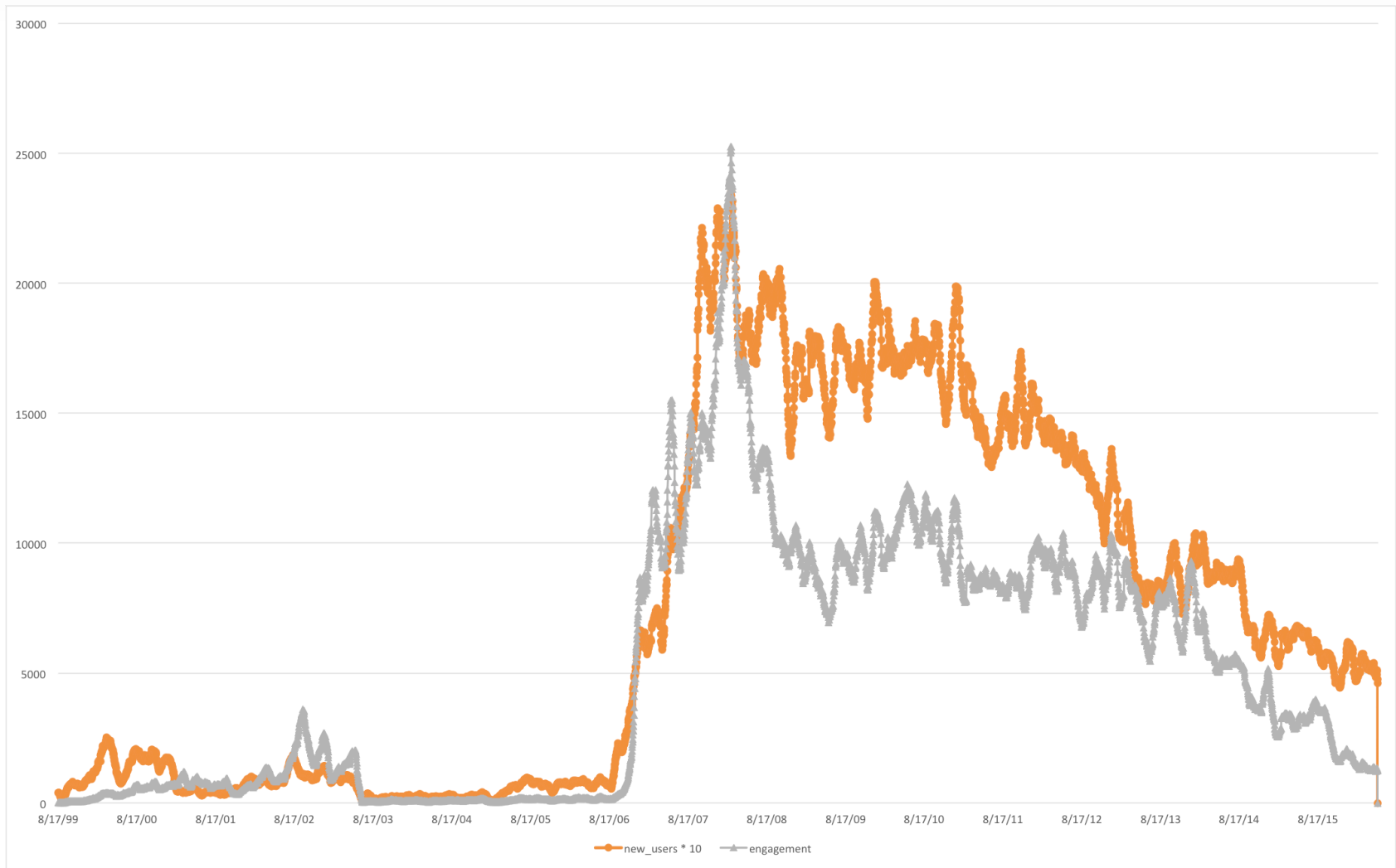
Analyze temporal, moving window-based, engagement capacity dynamics (e.g., per week); empirically validate the claim that higher engagement leads to higher reach

The objective is to gauge the correlation between the engagement value dynamics and:

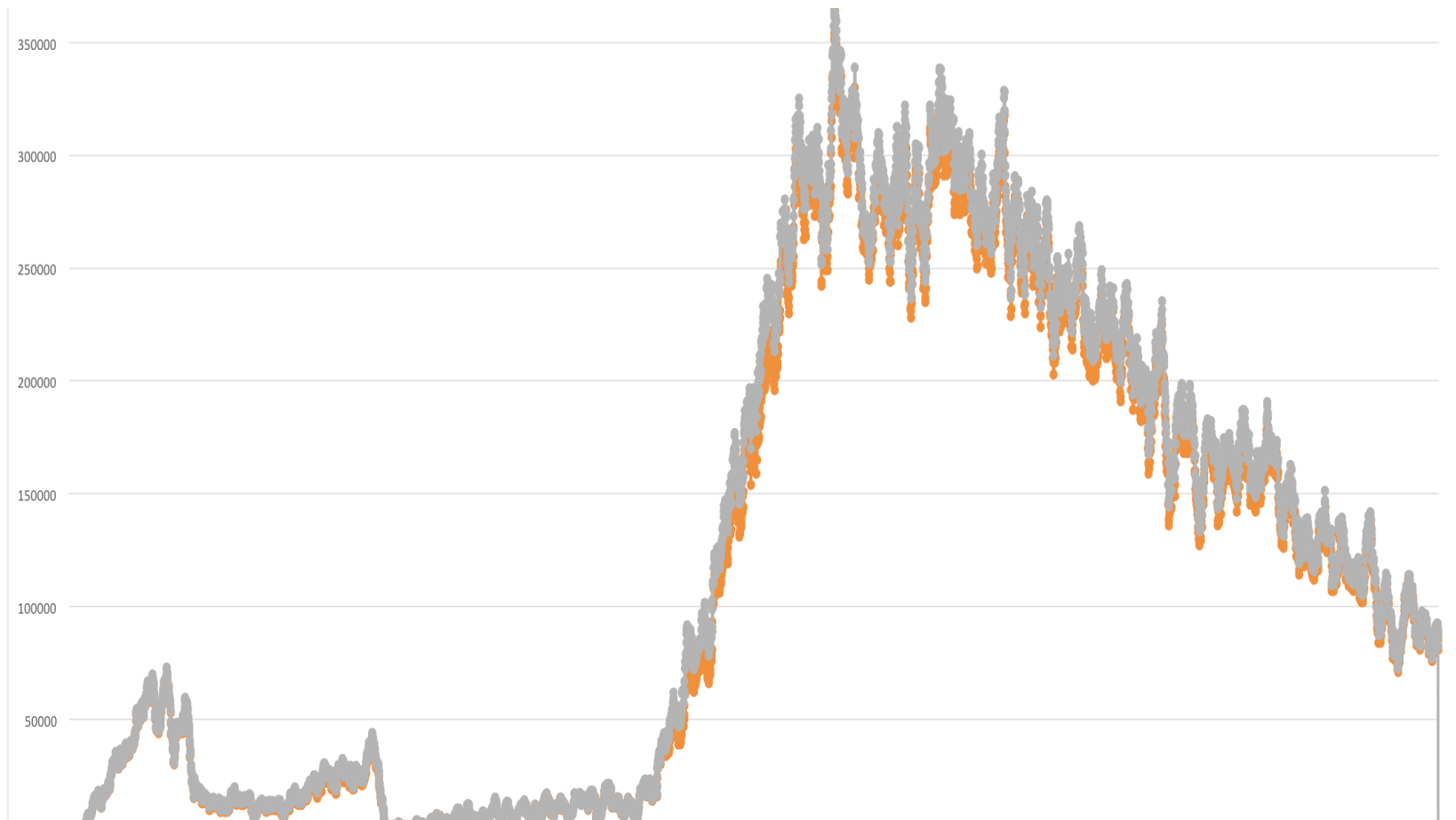
- addition of new users
- delurking



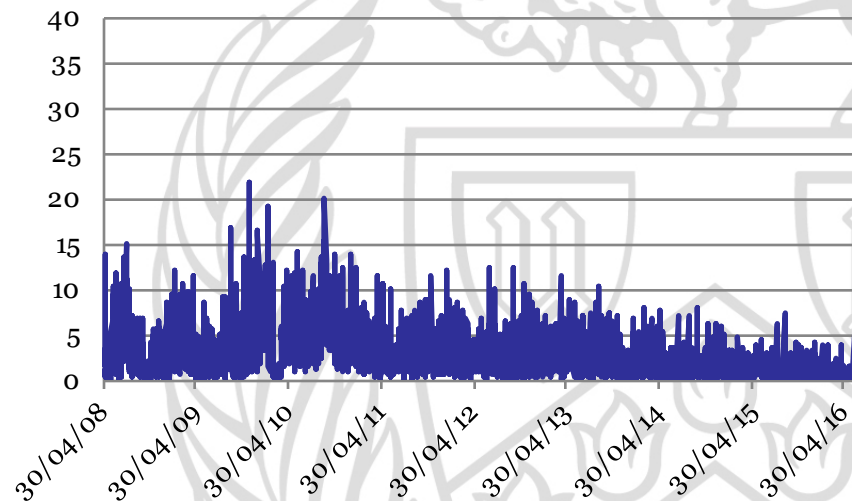
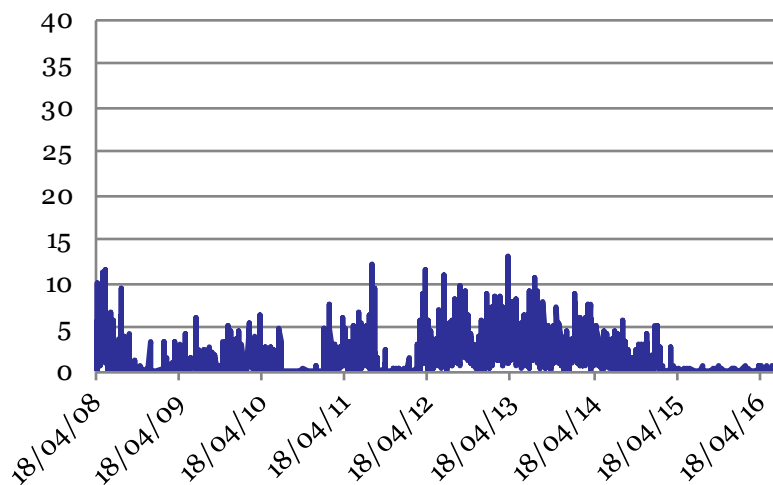
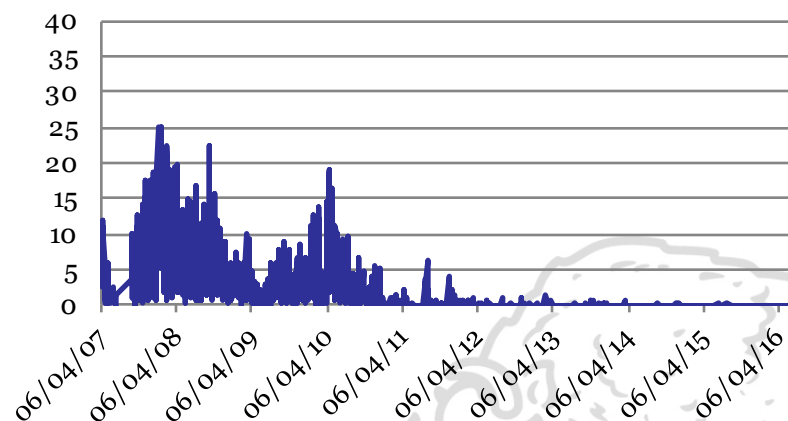
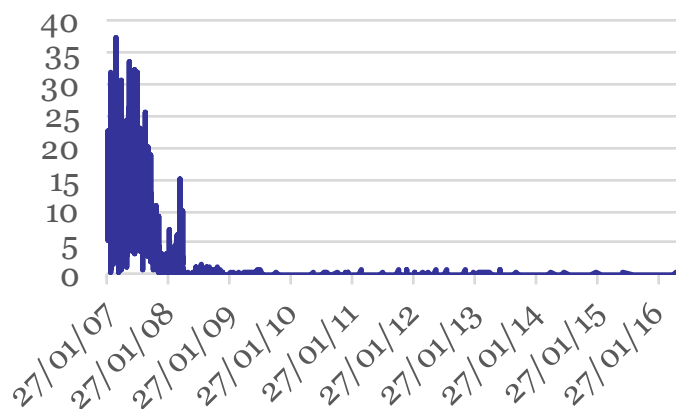
Engagement vs. New User Arrivals (x10)



Engagement vs. De-Lurking (x10000)



Engagement Capacity of the top users of the forum over time



Thanks,
Questions?

