Algorithmic Curation: Understanding the Content Network

Cody Baker (University of Evansville), Emilee Rader (MSU, BITLab)

Purpose

Sites like Facebook commonly use filtering algorithms to order and limit the amount of information their users see on a daily basis. We seek to understand the system level consequences of this, as well as it's impact on the behavior of individuals.

Approach

Agent-based modeling is a computational approach used to replicate a simpler, fully controllable version of reality. Such a model works by attributing values to the agents (in our case, people in a social network), assigning rules for how those agent interact with each other, and tracking how those interactions lead to emergent features within the system.


0	0.3	1

Clustering Coefficient
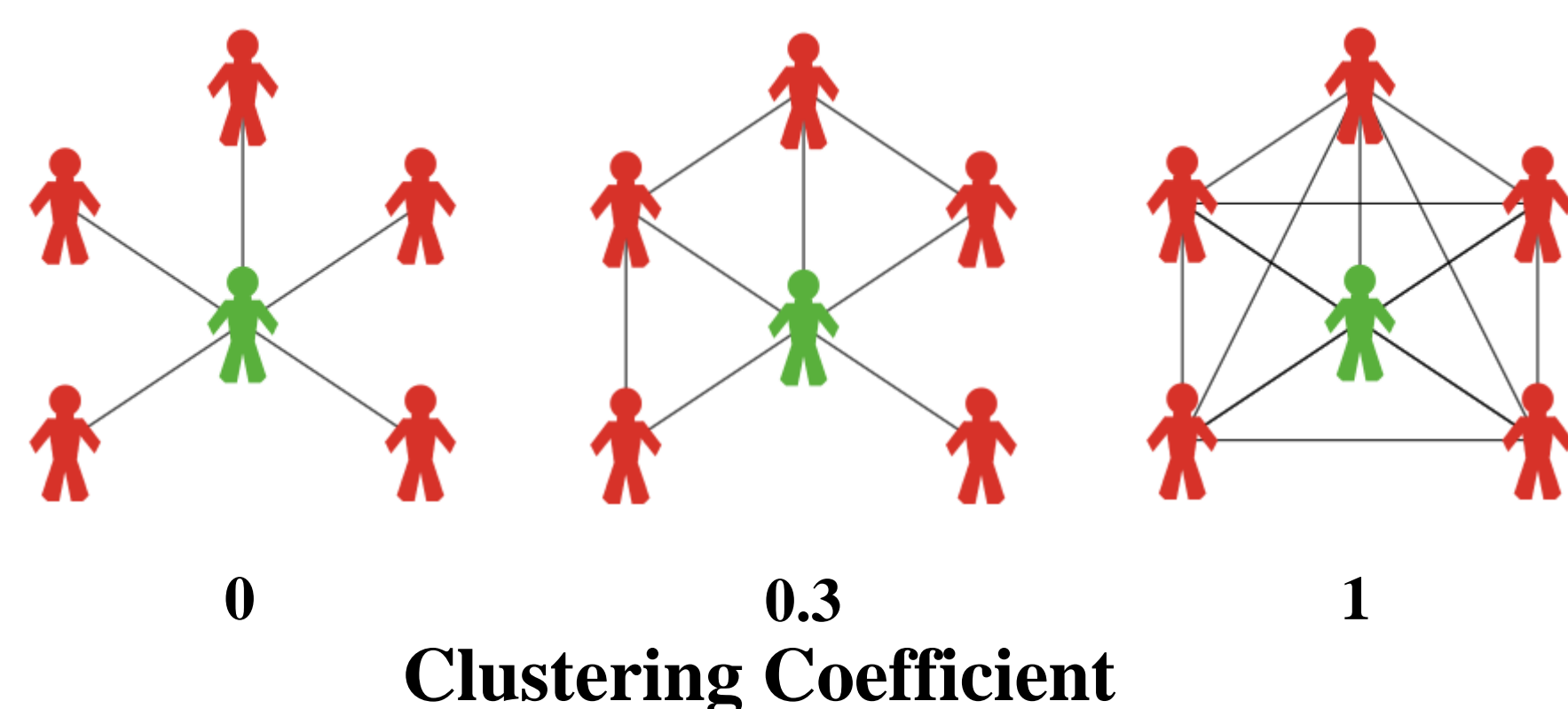
Modeling Facebook

With more than 1.28 billion users world-wide, a true replication of the Facebook network is computationally intractable. Thankfully, networks that exhibit the same statistical properties often behave very similarly system-wide, regardless of the scale. Thus, we created a much smaller version of such a social network and matched its parameters with that of the real Facebook.

Results

The people in the model engaged in posting and liking behaviors. Their posts were then filtered according to three conditions.
- Unfiltered (baseline)
- Distance (friends prioritized over friends-of-friends, etc.)
- Interest (measured on a popularity scale)
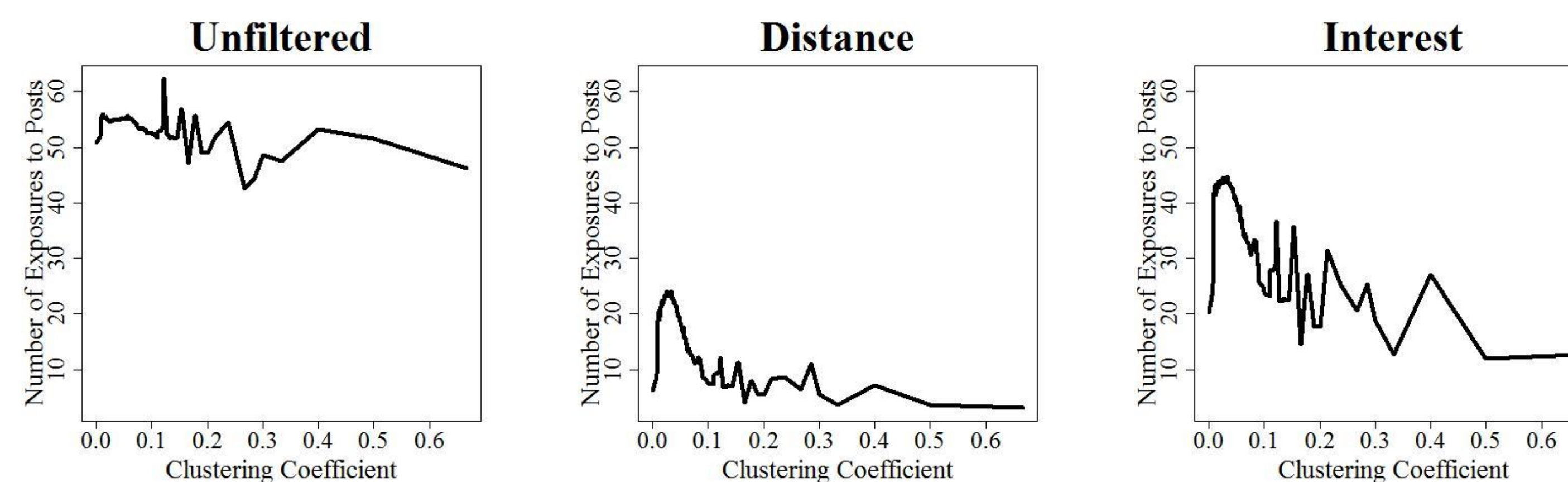
Within the network, some people tend to be more influential than others. We took the top percentile of these 'influencers', and tracked their behavior across the conditions.

Before Filtering   **After Filtering**
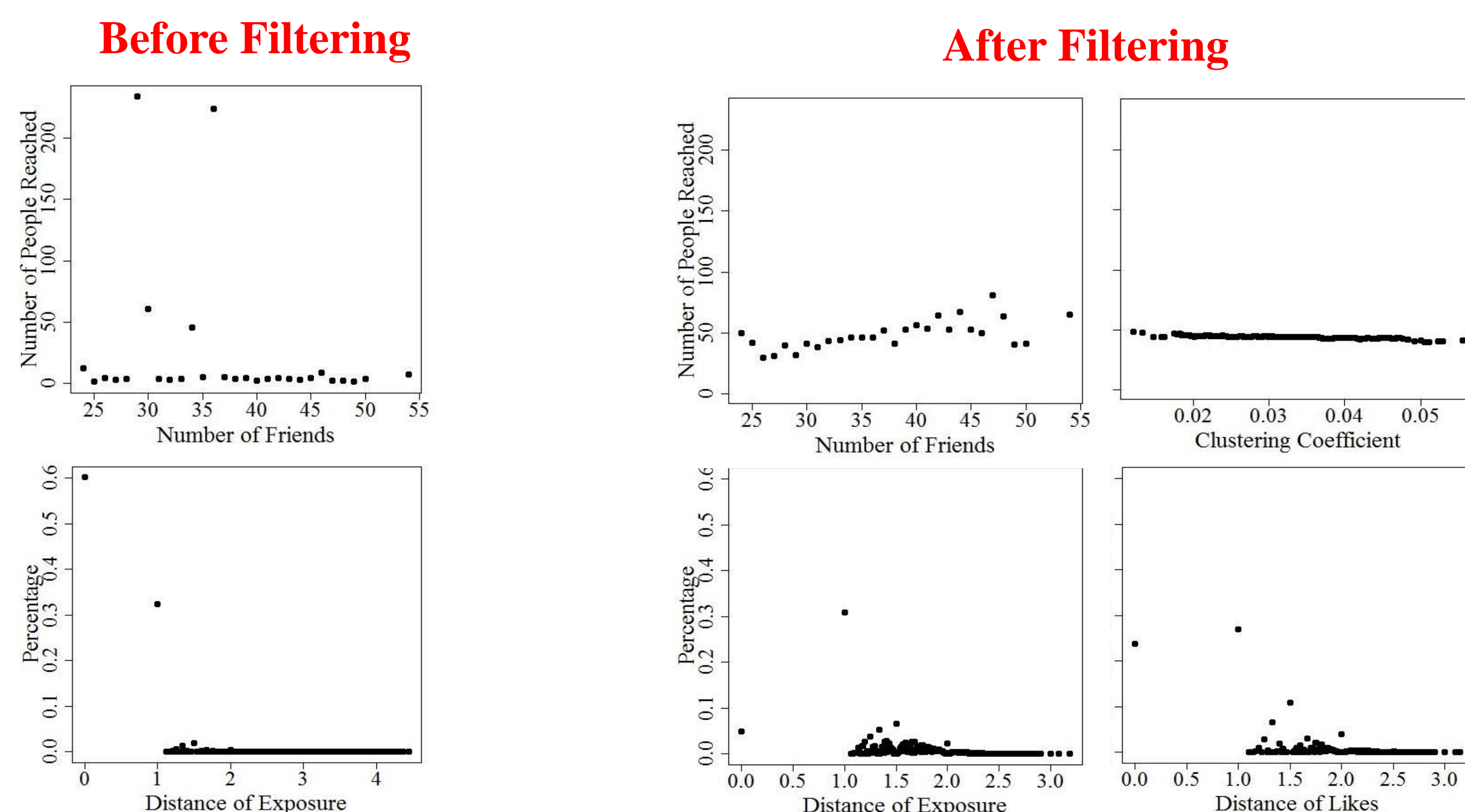
We also tracked the average distance a given post travels through the network, both in terms of that post being seen and 'liked'.

Number of Friends	Friend-of-friends	Degree of Separation	Clustering
Facebook			
Similarity 99%	87%	97%	90%
Model			

Conclusions

Individuals with an ideal clustering coefficient were exposed to significantly more content than others.

- This is likely because those individuals who possessed a friend network that was too disconnected (low cluster value) were presented with an overload of content, which resulted in excessive filtering.

- Whereas those individuals who possessed a friend network that was too well connected simply had less content to be filtered.

The amount of exposure of posts made by influential individuals was drastically increased by filtering.

Posts travelled much shorter distances after filtering, and assumed a different distribution over that interval.

Implications

Past research has primarily focused upon the behavior the content consumers rather than the producers.

The results of this study indicate that filtering may have a large effect not only upon what content an individual sees, but the content that is seen by users on a global scale.

System-wide exposure to posts by specific individuals can be easily controlled by the choice of filtering algorithm.