

# Online Influencers Not So Influential After All: A Critical Look At Who Popularizes Slang Online

Zachary Crenshaw

Department of Linguistics, University of Chicago

## Introduction

It is commonly thought that famous people (politicians, celebrities, activists etc.) are highly influential in defining new terms and introducing them to the public. For this reason, they are often referred to as "influencers". However, recent studies have found that influencers do not have this expected effect on social media (Zhang et al., *J Comput Mediat Commun* 21:1, 2016). According to this research, it is the "crowd", or the many ordinary users with few followers, that tend to have the largest influence on the popularity of social media trends. Prior research considered people to be influencers, and was conducted using social media sites focused on delivering content to users based on authorship, such as Facebook. This project tested social media that delivers content to users on the basis of the content itself, where posts can be considered influencers, not people. The social media site I chose is Reddit.com, a popular content-based forum website. Content is brought to the top of news feeds based on the popularity of the individual post, not by the popularity of the poster. The purpose of this study was to determine if influencers are leading or lagging indicators of new trends in slang usage.

## Methods

Reddit is composed of numerous forums, known as "subreddits." These subreddits represent distinct communities on the site, with users who subscribe to a subreddit constituting its membership. I chose five known, recent slang terms, picked to represent slang from a number of known subcommunities of English: *deadass*, *def/defo*, *periadt/periot*, *stan*, and, *thot*. Two of these terms have two variants, chosen to potentially find trends in variations of known slang. I then identified 10 subreddits that I expected would use the chosen terms at various frequencies: *AMA*, *askgaybros*, *askmen*, *asknyc*, *askreddit*, *askwomen*, *relationships*, *roastme*, *talesfromtechsupport*, *talesfromretail*, and, *tifu*.

I wrote Python scripts to utilize the PushShift API to scrape data on the frequency of use of these terms in these communities over the past 3 years. I also wrote scripts to find the influencers in the communities over the past three years. Influencers are defined as the top 1% of posts or comments in a subreddit in terms of score. Score is a popularity metric generated by members up-voting and down-voting a post. A higher score means a piece of content is more popular, and the content that is the most popular tends to be the most viewed, just as influencers have the most followers, and therefore views, on other social media.

The data was analyzed and visualized to find spikes in usage. A spike in content is defined as a 100% increase in usage over a period as large as 100 days with a minimum of 5 instances (posts, comments, or users). This minimum is designed to prevent spikes that occur from changes in low numbers of uses. This large increase threshold is designed to guard against small trends, from which it is more difficult to draw correlations with influencer spikes. A spike in influencers is defined as a 100% increase in usage over a period as large as 100 days, with no minimum in number. The data was collected on 10 day intervals and smoothed using a rolling average over 5 intervals, or 50 days.

## Results

Under the schema used, in the past three years there were 2 spikes in posts and 7 spikes in comments. There were some increases below the spike threshold, but in many cases the term usage remained fairly constant. The number of users publishing the term generally tracked the raw content numbers. This shows that spikes present in content were not the result of just a few users producing much content rich in these terms, but there was actually an increase in popularity of the term.

There were 84 post influencer spikes, and 15 comment influencer spikes under this schema. If the influencer spikes were held to the same minimum post threshold as content, then there were 10 post influencer spikes and 2 comment influencer spikes.

Figure 1

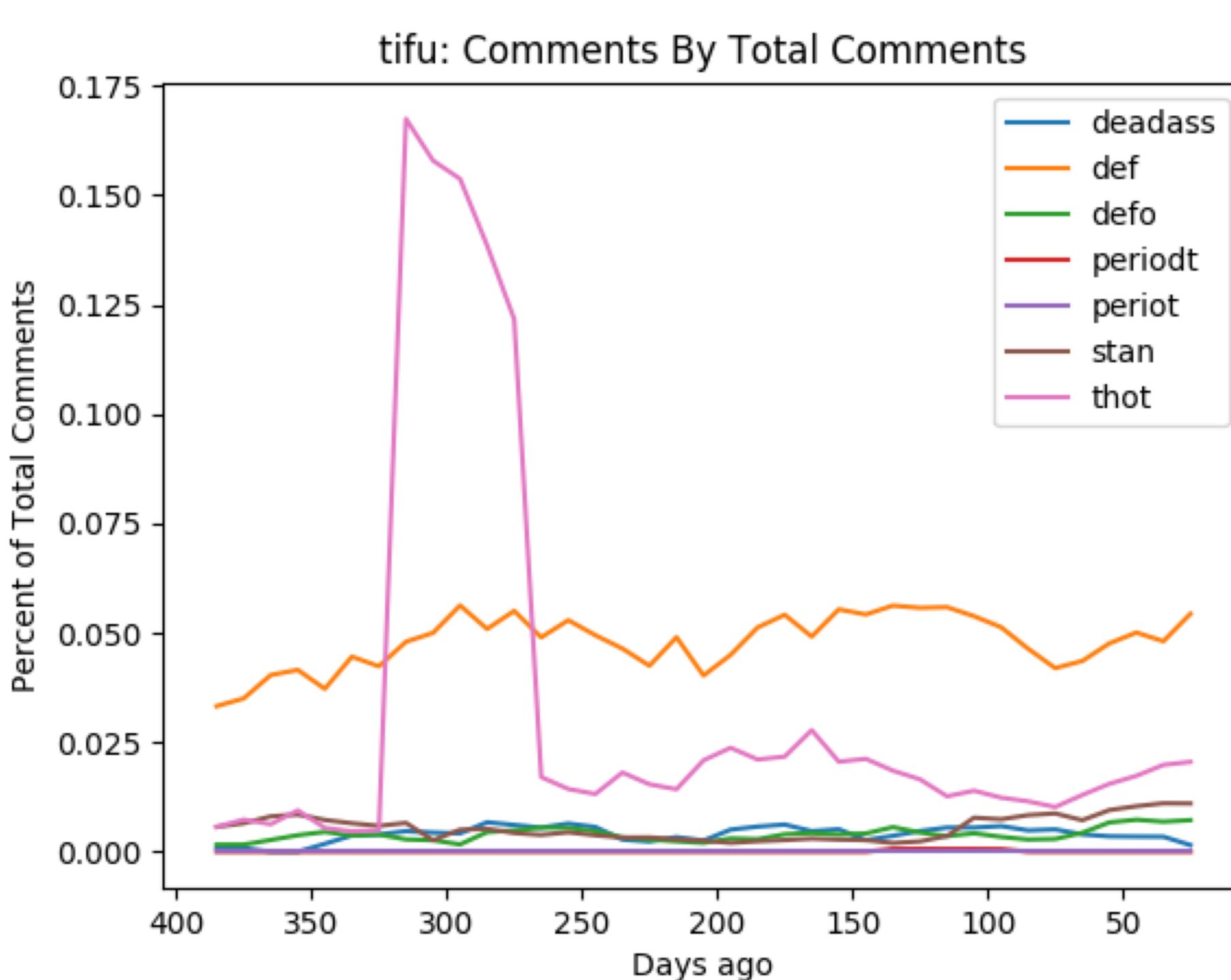
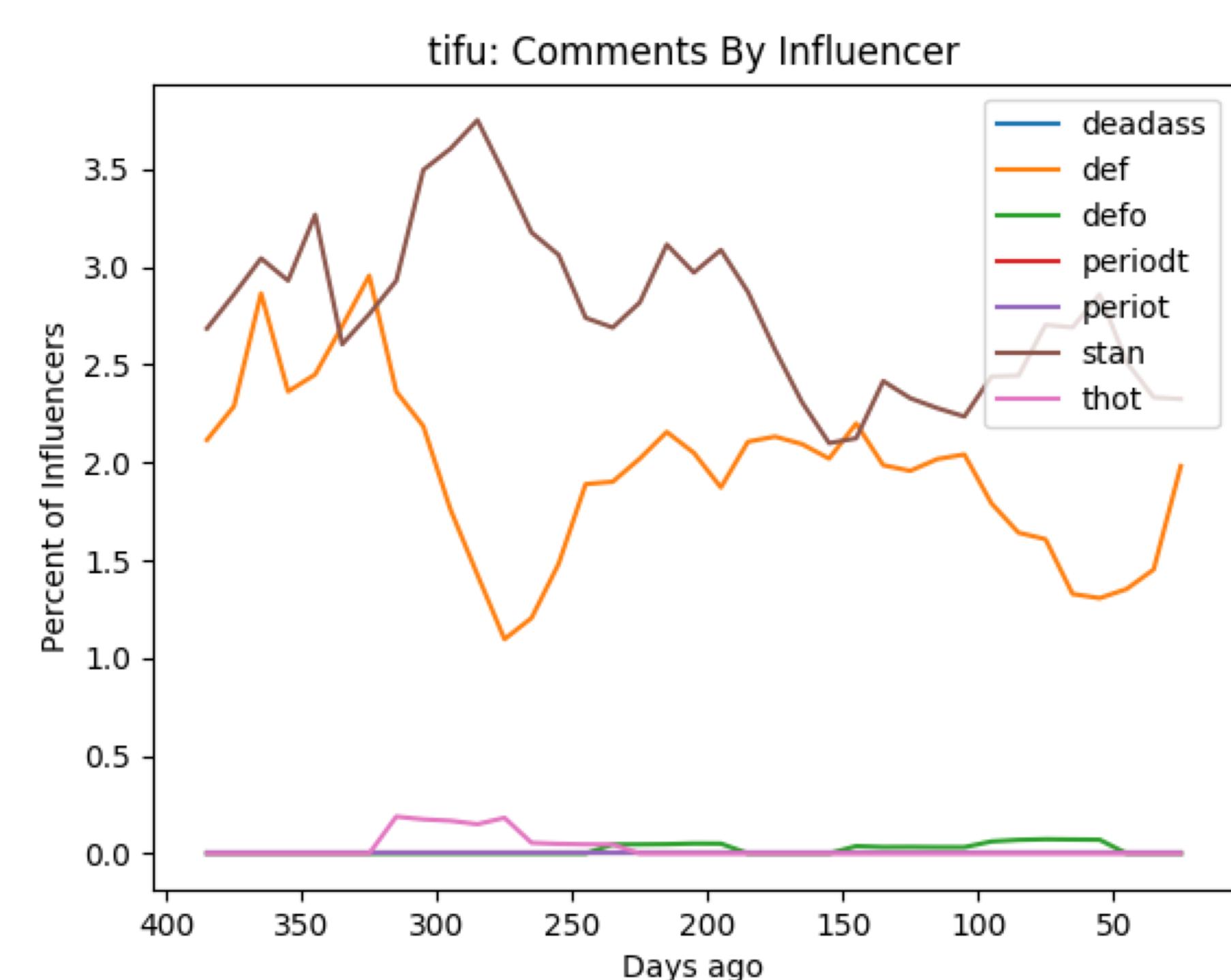
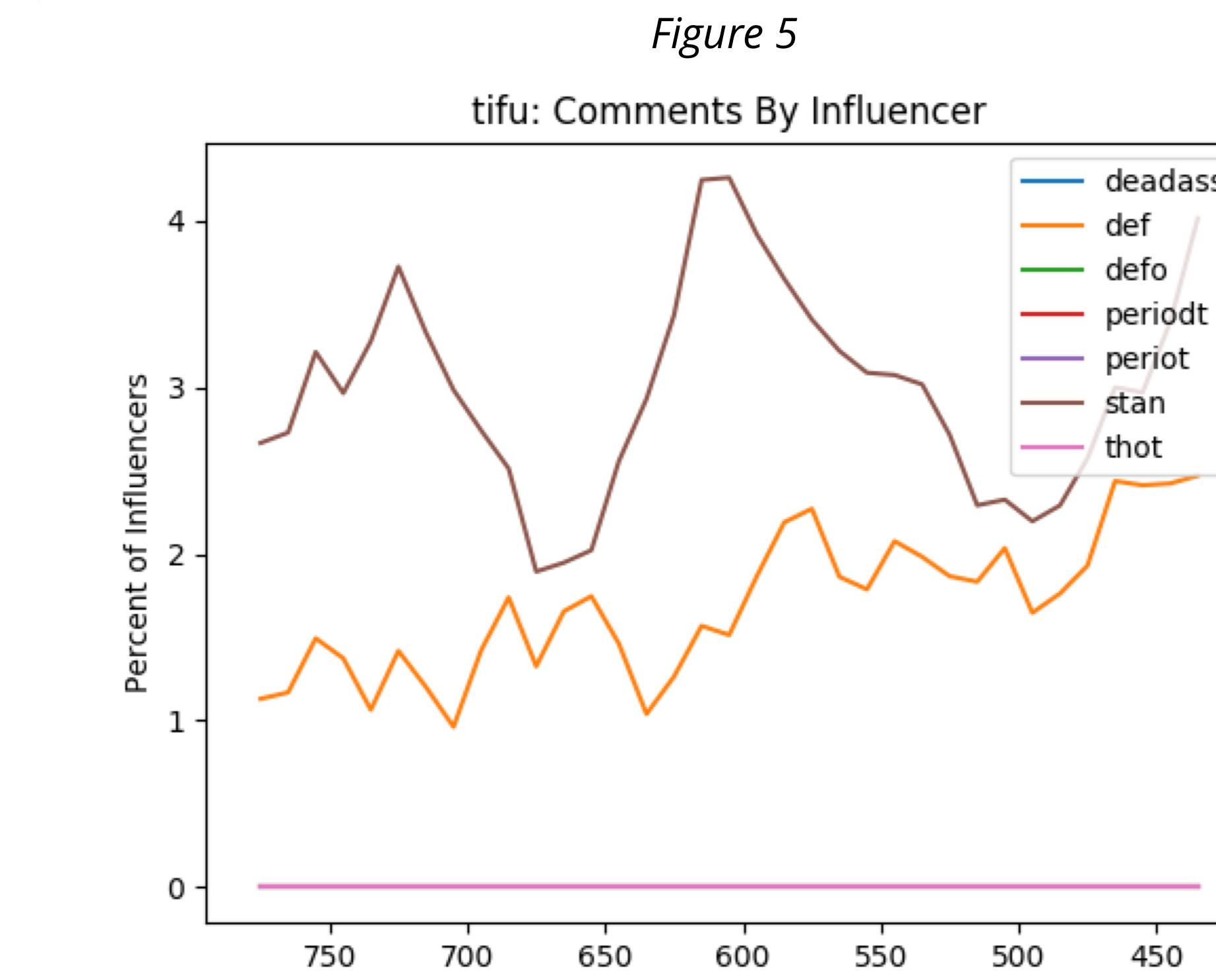
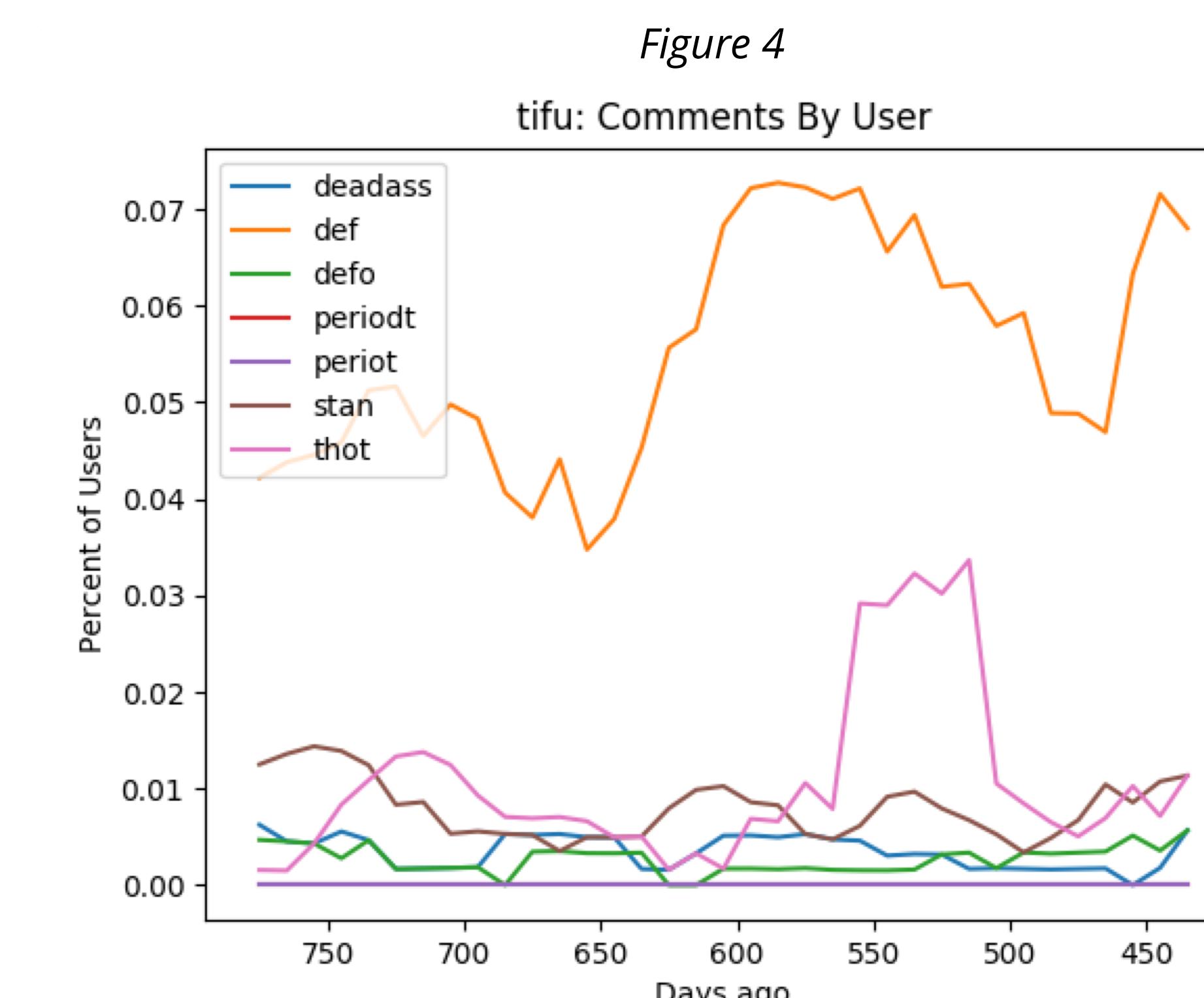
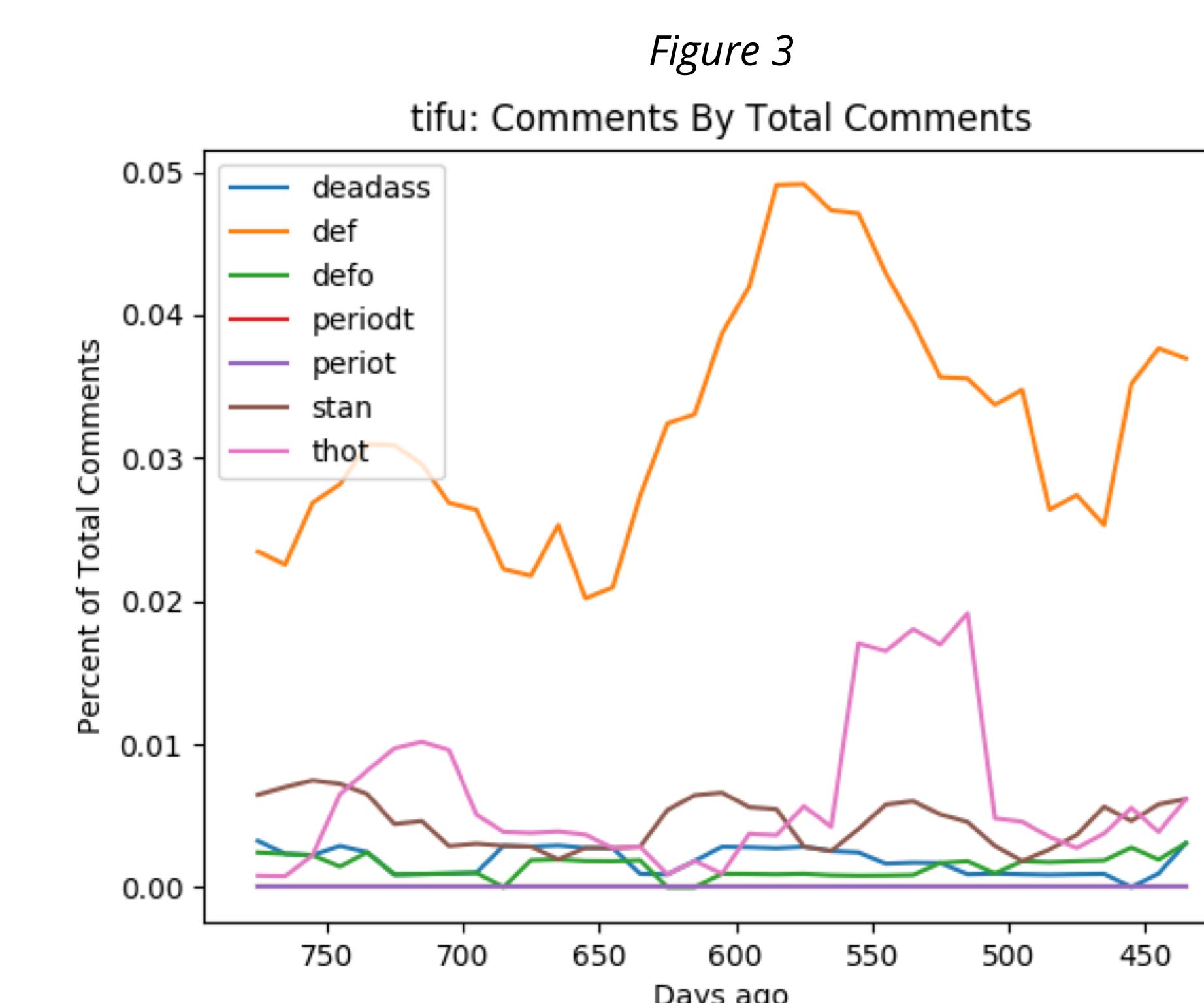


Figure 2



## Results



## Conclusion

The data does not support a clear relationship between influencer use of slang terms and crowd use of slang terms. While there were a few instances of influencers having a slight increase in usage around increases in crowd usage, these spikes were not substantive, nor were they consistently related to overall spikes, or even increases, in usage. When the influencer schema was modified to reduce apparent volatility, those spikes were not in the same time frame or for the same terms as the crowd spikes. These data suggest that there is little to no correlation between influencer usage and crowd usage. Furthermore, even when there was some correlation between influencer and crowd usage, the influencer posts constituted the majority of the usage, showing that the crowd did not follow or respond. Figures 1 and 2 show that, in some cases, there was a weak correlation to spikes, or simply increases, in both content and influencer usage rates. Figures 3, 4, and 5 are examples of the more common case that the content and user rates correlate, but the influencer rates do not strongly correlate to the other two.

## Discussion

While in this schema influencers seem to have little influence on slang, there are some areas for improvement on the analysis. The definition of spikes could be refined further to include smaller spikes, or hills (more gradual increases in usage). These were excluded to limit the scope of the project to challenge the idea of a strong correlation, but a weaker one may exist. A more robust schema to reduce volatility in the data could be implemented. Part of the volatility is likely due to the small number of influencer posts. This could be mitigated by redefining the term influencer to be more inclusive of lower ranked posts, especially for small subreddits. However, as the percentage of posts that are influencers is increased, the classical notion of influencers as being a few very famous entities breaks down. As this notion breaks down, the line between influencer and crowd begins to blur, making it difficult to properly test the hypothesis.

## Acknowledgments and Citations

I would like to thank Jason Riggle, Assistant Professor, Department of Linguistics, University of Chicago. Not only did he help sponsor this project for funding, but he has also been a wonderfully insightful mentor.

I would also like to thank the University of Chicago Dean's Scholarship program for funding this research to make it possible.

Citations:

Leihan Zhang, Jichang Zhao, Ke Xu, Who creates Trends in Online Social Media: The Crowd or Opinion Leaders?, *Journal of Computer-Mediated Communication*, Volume 21, Issue 1, 1 January 2016, Pages 1-16, <https://doi.org/10.1111/jcc4.12145>