

Posts on Central Websites Need Less Originality to be Noticed

Keywords: Social media, viral information, network science, memes, reddit

Extended Abstract

The Internet has created information highways through society. Via social media, news feeds, newsgroups and instant messaging, people are able to spread messages with an unprecedented speed and reach. These capabilities have been a force for good, but also for bad. There is a wealth of research showing that misinformation spreads more rapidly [1] and in larger cascades [2] on social media than true information. In general, it is imperative to understand better how and why messages go viral, to encourage the positive effects of the Web and mitigate its potential negative ones.

One factor that seems to be obviously related to the likelihood of viral diffusion is the content of a message. However, the role of content in determining virality is currently explored mostly qualitatively, with a focus on perceived rather than objective characteristics. To objectively quantify the relevant aspects of a message's content is hard. Simple quantitative measures of content are typically found not to predict success, or to be overwhelmed by other factors [3].

A few papers, however, explore more sophisticated measures of content and find them to be relevant for message success. Factors range from the sentimental valence of the message, the amount of arousal it provokes, and originality – the latter seems to help posts be noticed, even if it does not help with success per se [4].

Parallel to the study of successful content, research has also looked at the effects of para-textual elements on success: how quickly appreciation from users flows after content publication, rich-get-richer effects for messages that obtain an early advantage, etc. One para-textual element that has received particular attention is source centrality: whether the originators of a message or the websites they use to publish their content sit at the intersection of many communities or are otherwise central in the social structure [5].

The success-enhancing capacity of network hubs comes partly from their structural ability to reach many other nodes [6]. But hubs also serve as a brand name for the content they host. For instance, a Reddit user may have an easier time getting upvotes for a post with a YouTube video than for one with a video hosted on their personal website. Because of their reach, hubs (such as YouTube) are more likely to be familiar, and the so-called recognition heuristic should drive users to attach more value to their content [7]. At the same time, the recognition heuristic plays a smaller role when other quality cues are available – e.g., when the content itself is highly original. In the case of social media, most sites display content more prominently than source, meaning that original content may be eye-catching enough for users not to rely on the recognition heuristic.

In this study, we explore the possibility that such interactions between content and para-text are part of the reason why some studies have found surprisingly small effects of content on success. Specifically, we propose that content hosted on websites that are central in the Web, and therefore highly recognizable, need less originality to be receive attention. Put differently, originality has different effects depending on the level of prominence of the information source. We measure originality by averaging the bigram conditional probability of the words in the

title, while centrality is the in-degree of the website publishing the content in the Web network – results are robust to different originality and centrality measures.

Our results indicate that originality interacts with centrality in predicting the odds of a message getting noticed at all. Originality matters more when the message originates from a peripheral website. Central platforms can get away with less original messages, most likely because their recognizability helps them get at least minimal attention. Figure 1(a) shows that the probability of getting more than one upvote on Reddit grows with our originality measure, but only for low centrality sources. For high centrality sources this probability actually decreases.

On the other hand, the effect of this interaction is only clear when estimating the chances for a message of not failing – on Reddit, this means getting at least one upvote, which represents the minimum possible bar of attention. The effect is unclear when looking at the chances of becoming one of the most successful posts. Figure 1(b) shows that originality is not a good predictor of the probability of being among the top 10% performing posts in a day on Reddit – in number of upvotes. This is regardless of centrality.

References

- [1] David MJ Lazer, Matthew A Baum, Yochai Benkler, Adam J Berinsky, Kelly M Greenhill, Filippo Menczer, Miriam J Metzger, Brendan Nyhan, Gordon Pennycook, David Rothschild, et al. The science of fake news. *Science*, 359(6380):1094–1096, 2018.
- [2] Jonas L Juul and Johan Ugander. Comparing information diffusion mechanisms by matching on cascade size. *Proceedings of the National Academy of Sciences*, 118(46):e2100786118, 2021.
- [3] Justin Cheng, Lada Adamic, P Alex Dow, Jon Michael Kleinberg, and Jure Leskovec. Can cascades be predicted? In *Proceedings of the 23rd international conference on World wide web*, pages 925–936, 2014.
- [4] Michele Coscia. Average is boring: How similarity kills a meme’s success. *Scientific reports*, 4(1):6477, 2014.
- [5] Lilian Weng, Filippo Menczer, and Yong-Yeol Ahn. Virality prediction and community structure in social networks. *Scientific reports*, 3(1):1–6, 2013.
- [6] Javier Borondo, Florentino Borondo, Carlos Rodríguez-Sickert, and Cesar A Hidalgo. To each according to its degree: The meritocracy and topocracy of embedded markets. *Scientific reports*, 4(1):1–7, 2014.
- [7] Gerd Gigerenzer and Daniel G Goldstein. The recognition heuristic: A decade of research. *Judgment and Decision Making*, 6(1):100–121, 2011.

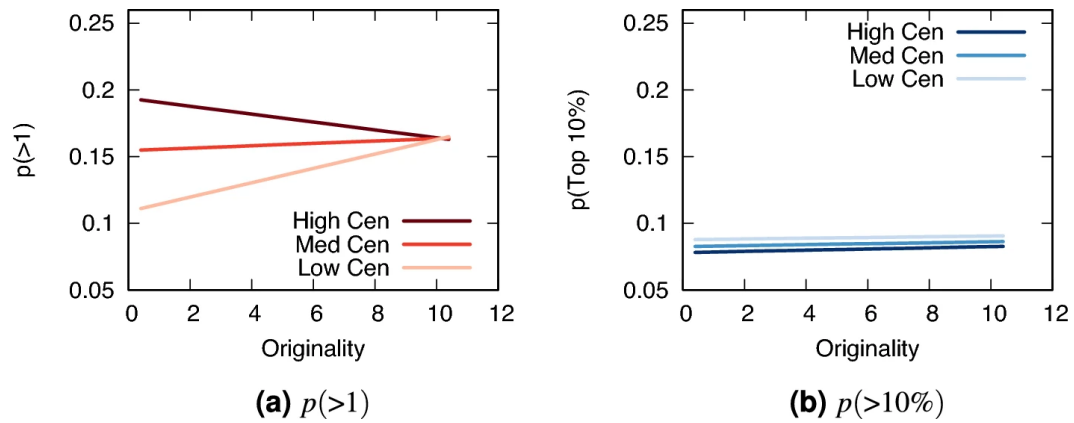


Figure 1: The interaction effect between originality and centrality. Originality on the x axis, centrality (“Cen” in the legend) is the darkness of the line, success on the y axis. (a) Probability of getting at least one upvote ($p(> 1)$) in red; (b) probability of being in the top 10% scoring posts in a day ($p(> 10\%)$) in blue.