

# Estimating Diffusion of Music on an Online Social Network

---

Rajiv Garg, David Krackhardt, Mike Smith, Rahul Telang

## ABSTRACT

Peer influence in social networks has been studied for over four decades by social scientists. But it is only recently that it has gained attention from computer scientists and online product managers. But most extant empirical work measuring peer effect is particularly challenging due to selection problem, separation of influence from other peers, and accounting for a user's pre-existing knowledge. In this paper we overcome these challenges and find that the online peers have a small but statistically significant influence on other peers and enable discovery of music on online social networks. Thus, we cleanly and empirically estimate that online users are 6 times more likely to discover a new music because of their peers in an online social network.

## RESEARCH QUESTION

In this research we examine if peers diffuse music to other users, and whether and how big a role social networks play in users' discovering new content. The goal is to cleanly identify the extent of diffusion while addressing various challenges associated with estimation of peer influence.

## APPROACH

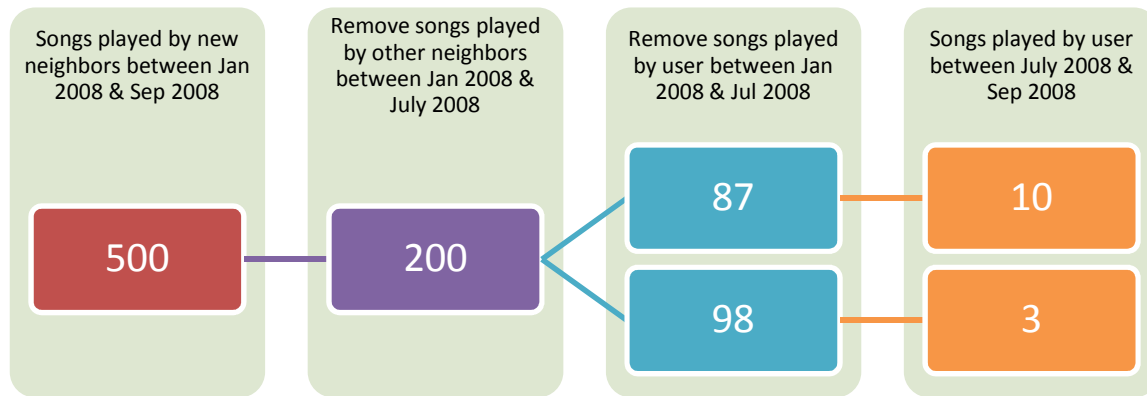
We collected a large volume of data (approximately 99 million data points) from the online social network established on Last.FM. The goal is to utilize that publicly available dataset to address the issues of user selection, influence separation, and existing knowledgebase for estimating the role of online peers on diffusing a new piece of music.

We pick "neighbors" (users recommended by Last.fm system) as peers for source of diffusion. These neighbors (recommended peers) are mostly strangers to a user and are recommended based on their matched interest in music. Thus these peers have no other mode of communication with the users except for the Last.fm network itself. Additionally we screen all the content made available by these neighbors to filter all songs already played by anyone else in the network of the user. Thus together the filter and the neighbors as peers emulate (to some extent) the closed environment where there is no external influence on a user's discovery.

To screen the pre-existing knowledge of a user, we remove all the songs that were heard by a user in the previous timeframe. Thus the pool of songs that could be diffused to a user by the neighbors includes only those songs that are neither played by the user nor anyone else in the user's neighbor network. This was enabled because of the large volume of archival data collected from Last.FM , which included all the songs played by users in the previous 3 years.

Finally, we pick control users to account for any by chance discovery of songs available for diffusion. This control user population is similar to the users discovering the songs, and is not connected to the neighbors diffusing the songs. These set of control users then allow us to estimate the discovery of new content from sources other than the diffusing neighbor and thus establish a baseline for our peer influence estimate.

The figure below presents our approach in estimating the peer influence.



## RESULTS

In this paper, we find that there is a positive influence of online peers on diffusion of new music. Users are 6.1 times more likely to discover a new track and 3.4 times more likely to discover a new band as a result of peer influence. A key contribution of our paper is: (i) we provide a clean test for diffusion on online networks and are able to overcome many key challenges in estimating peer effects. Moreover, we do this in a field setting without resorting to survey or laboratory setting. Thus our paper provides a roadmap for how to use large noisy data and still be able to identify the peer effects with reasonable confidence, (ii) we provide an empirical evidence of how even passive networks where the peers do not know one another can help in information discovery. Even new songs seem to diffuse on these networks. This suggests a significant power of online networks in content discovery.

## STATUS

The earlier version of this paper has completed the estimation of peer influence and the role of user's listening behavior on the extent of new music discovery. We have furthered the findings by incorporating the results on similarities between the treatment and control group of users. Currently we are planning to collect survey data from the studies social network on Last.fm to complement our findings. We assume that the results from a random group of users will allow us to evaluate user's behavior towards new music and their online peers. Thus we would like to extend and present our results by studying diffusion from contrasting studies – analysis of archival data and inference deduction from survey data.