Title: Diversity and Networks in Artificial Cultural Markets

Author: Kevin Lin

Department: Department of Computer Science and Information Science

Faculty Advisor: Michael Macy, Ph.D., Department of Sociology and Information Science

Abstract: For every hit song, movie, and TV show, there are hundreds more that are flukes. This pattern of market domination by the most popular subset of cultural items demonstrates that they must have a greater intrinsic value. However, experts routinely fail to predict the success of these cultural items, suggesting that the intrinsic value of the most successful products is not so easy to predict. The Music Lab project (Salganik, Dodds, & Watts, 2016) sought to understand this phenomenon, by creating artificial music markets in multiple worlds. The participants in the study were randomly assigned to a world and downloaded from a set of unknown music either with or without the knowledge of previous participants’ choices. The study showed that social influence increased both the inequality and the unpredictability of success, causing people to diverge from the choice of greatest intrinsic value. The purpose of my summer project will be to refute the study, and to show that in the specific case of the Music Lab project, after a long enough period of time, the popular song will converge back to the song with greatest intrinsic value. Then, I will run simulations to seek to understand, in a more general scope, under what conditions do the popular choice of a given world converge back to the product of greatest intrinsic value.
Biographical Sketch

Born in Colorado, Kevin lived in a small, cozy town by the mountains for three years before moving to Maryland. He has lived in quite a few cities along the east coast, but currently calls Houston, Texas home.

After the stint in Houston for high school, he decided to head back east for college. He loves attending Cornell for its academic rigor, the beautiful weather, and the vibrant community of people.

Kevin is a sophomore in the College of Arts and Sciences majoring in Computer Science. His main interests are understanding human behavior and interactions through a technical lens. He enjoys using computational techniques to analyze trends in human tendencies and behavior on a large scale.

Some of Kevin’s favorite experiences at Cornell involve the activities and classes outside of his major. He was never a dancer, but decided to join a few dance clubs on a whim. Now, some of his best friends are the ones he made there. Kevin is also grateful for the wonderful language-learning opportunities at Cornell. After polishing up his Spanish, he decided to begin learning Russian. Despite the daily classes and time commitment required, Kevin finds the process very fulfilling. Learning a language is learning a culture—how people think, behave, and understand.

At Cornell, Kevin is the president of the Swing Dance Club and the vice president of the Ukulele Club. He enjoys eating watermelon, playing ukulele, and dancing Argentine Tango, although not always at the same time.
Statement of Purpose

For every hit song, movie, and TV show, there are hundreds more that are flukes. This pattern of market domination by the most popular subset of cultural items demonstrates that they must have a greater intrinsic value. However, experts routinely fail to predict the success of these cultural items, suggesting that the intrinsic value of the most successful products is not so easy to predict.

The Music Lab project (Salganik, Dodds, & Watts, 2016) sought to understand this phenomenon, by creating artificial music markets in multiple worlds. The participants in the study were randomly assigned to a world and downloaded from a set of unknown music either with or without the knowledge of previous participants’ choices. The study showed that social influence increased both the inequality and the unpredictability of success, causing people to diverge from the choice of greatest intrinsic value.

The results of this study are indeed very important because it provides empirical evidence that in the presence of an information cascade, there will be a divergence of opinion among the multiple worlds. In each independent world, by random chance, a different opinion can arise as the most popular one locally, and because of the effect of social influence, that local opinion will propagate onwards in the form of an information cascade, despite what is actually enjoyed and objectively preferred (Salganik & Watts, 2008). Imagine a new song comes out that you don’t particularly like. However, enough of your friends have already expressed interest in the song, that you buy it anyways. When this case is generalized, depending on who happens to like a song at the earliest time, each world could end up with drastically different outcomes.

The purpose of the research conducted this summer will be to explore this phenomenon within multiple worlds in greater detail. One of the main complications that arises from the Music Lab study is the following: If an information cascade can lead multiple worlds to diverge from the objectively best option, if sufficient time has passed, would the worlds eventually converge back to the option of best intrinsic value? The rationale behind this question is intuitive. If there is a process causing people to stray further and further away from their own preferences, is there at the same time another process that slowly brings people back towards them?

I will attempt to refute the claim in the Music Lab study, and show that in that specific case, it is true that after sufficient time has passed, the consensuses within each world do eventually converge back to the objectively best option. This can be done by analyzing the data published by the study, but on a longer time horizon. Because there was a world that acted as a control, in which there was no social influence, I hope to find that given a long enough time, the consensuses of each of the experimental worlds will converge back to the consensus of the control world.
Furthermore, I would like to explore in greater detail the conditions that cause the convergence back to the objectively best option. The method of experimentation in this case would be simulation, and therefore, it would not require any human subjects. By carefully making assumptions about the peoples’ behavior, we can test various hypotheses given the assumptions.

The first hypothesis to test is the hypothesis of diversity in preferences. Obviously, in the real world, not everyone has the same preferences, but it is unclear the effect of the diversity of preferences on the convergence of opinions to the control. The hypothesis I propose is that in the case in which there is no diversity, that everyone prefers every item the same amount, there is no convergence because no one has a differing opinion to change the course of the information cascade back to the option of greatest intrinsic value.

At the other extreme, I propose that when the diversity of preferences is too great, then people will prefer options no better than random chance, and this will also fail to converge back to the option of greatest intrinsic value. Even though people are more willing to stray away from the information cascade, the same process also distances themselves from the objective control. Therefore, the diversity must be somewhere in the middle, at neither extreme, for the convergence. Using these simulations, I hope to learn more about the property of diversity, such as which distributions of diversity are required, and for what values is convergence the most probable.

The second main hypothesis to test involves the network dynamics within the worlds. In the prior hypothesis, it is indeed true that we are manipulating the preferences of individuals, because it is sensible that not everyone has the same preferences. It is also sensible that not everyone always picks their most preferred item, because the line between two preferred items is not always very clear.

However, one aspect of the experiment that is neglected is the network of people within the cascade. It is a large assumption to make that everyone is influenced by everyone else the same amount. The degree to which a person will accept the opinion of another greatly depends on a variety of factors. For example, people who are similar across a few demographic traits are more likely to agree on something. As such, the degree of homophily within the network certainly will influence how the information cascade changes course. Another parameter that has potential to influence the convergence to the objective best option would be the network structure. The network structure encompasses properties such as the connectivity of individuals, the degree of groupings, and concentration of individuals within the network. These all play a subtle but important role as to how social influence propagates through the network.

By understanding how these parameters affect the path of popularity in the multiple worlds, we would be able to better understand how the diversity of preferences and network dynamics affect the path of popular items within an simulated artificial cultural market.
Conducting the experiment by simulation has a few advantages over the traditional experimental approach. First of all, recruiting participants is expensive, especially in this case, because we need a sufficient number of people in each world over a long period of time. It is also difficult to learn attributes about the participants such as demographic features or network structure because they are random participants through the internet. People are for the most part unwilling to give too much information about these attributes away.

The simulated approach, on the other hand, makes up for the expensive recruitment and ambiguous attributes because all of the characteristics about the participants are computationally generated based off of assumptions. There is no limitation on the number or available information of the participants.

A possible drawback of the computational simulation is the validity of the conclusions that arise. However, I contend that it is nearly impossible to construct this version of the experiment in the real world because the underlying diversity of preferences and network dynamics are very difficult to ascertain. By careful construction of assumptions against previous research, the simulation can still lead to very interesting and subtle results, as long as we’re able to accept the assumptions are true.

For the purposes of conducting this research, I plan on staying in Ithaca over the summer to collaborate frequently with my advisor and the other members of the Social Dynamics Lab. I also tentatively plan on attending one or a few conferences on the topic of networks and computational sociology.
Bibliography
