Spreading the Oprah Effect: The Diffusion of Demand Shocks in an Online Recommendation Network

Eyal Carmi *(Tel-Aviv University)*,
Gal Oestreicher-Singer *(Tel-Aviv University)* and
Arun Sundararajan *(NYU Stern School of Business)*

Overview and Data: We study the rate and persistence of the diffusion of exogenous demand shocks through the co-purchase recommendation network on Amazon.com. Our goal is to understand whether such online networks form the basis for the contagion of exogenous economic events, and if so, what the *magnitude* of this influence/contagion is, how *far* the shocks "spread", how long they *persist*, how the magnitude and persistence of the shock is affected by the *clustering* and local structure of the underlying recommendation network, and under what conditions the shock leads to a *fundamental shift* in the demand curve for the product rather than simply being a fad-like shift.

Our data set is constructed by tracking book reviews in a popular television show (Oprah) and a popular newspaper (the New York Times) over the year 2007. We combine this information about exogenous events with existing data about the co-purchase recommendation network that originates from each of these reviewed books on Amazon.com, and the demand for each reviewed book as well as all of each reviewed book's first, second, third and fourth neighbors in the recommendation network over a six week interval, beginning two weeks before the event and ending four weeks after.

Since there are many paths by which one might reach a neighboring book over the recommendation network, in a sense, books with a larger number of paths terminating at them are "closer" to the reviewed books; besides, many books are both first neighbors and second neighbors. To normalize our measure of how "far" the shock diffuses, we construct a normalized graph distance measure that is inversely proportionate to the probability of reaching the book by randomly traversing the recommendation network starting from the reviewed book. We bin neighbors into "first neighbors", "second neighbors" and so on where the average distance of each bin corresponds approximately to being one click away, two clicks away, three clicks away and four clicks away, so that each neighbor has a unique distance from the reviewed book. Our data comprises a total of 53 reviewed books, 248 first neighbors, 665 second neighbors, 1543 third neighbors and 3444 fourth neighbors.

For each of these reviewed and neighboring books, we next construct a number of measures that measure the extent and persistence of the diffusion. A comprehensive list is beyond the scope of this abstract. Two important measures are the *demand inflation factor* (the ratio of the demand of the book on day t after the event and the average demand for the same book in the period trailing the event), and the *persistency duration* (the number of days after the event that it takes for the demand for a book to return to within one standard deviation above its average per-event demand – this is a very conservative measure of shock persistency duration on two accounts). Other measures we examine include the area under the diffusion curve, the aftershock ratio and the fraction of the shock passed on to immediate neighbors.

Preliminary Results: Our preliminary results provide compelling evidence that the exogenous demand shocks cause statistically and economically significant changes in the demand for neighboring books, that the shocks travel quite far in the network, and highlight the range of persistence of these shocks as well as their connection to the structure of the underlying recommendation networks.

Magnitude of diffusion: We find that on average, the first neighbors and second neighbors experience a demand shock that is statistically significant immediately following the event. Figure 1 illustrates the demand inflation factor for a candidate reviewed book along with candidate first, second, third and fourth neighbors While third and fourth neighbors do not show a statistically significant change in demand on average, a considerable fraction of these books do experience a statistically significant shock.

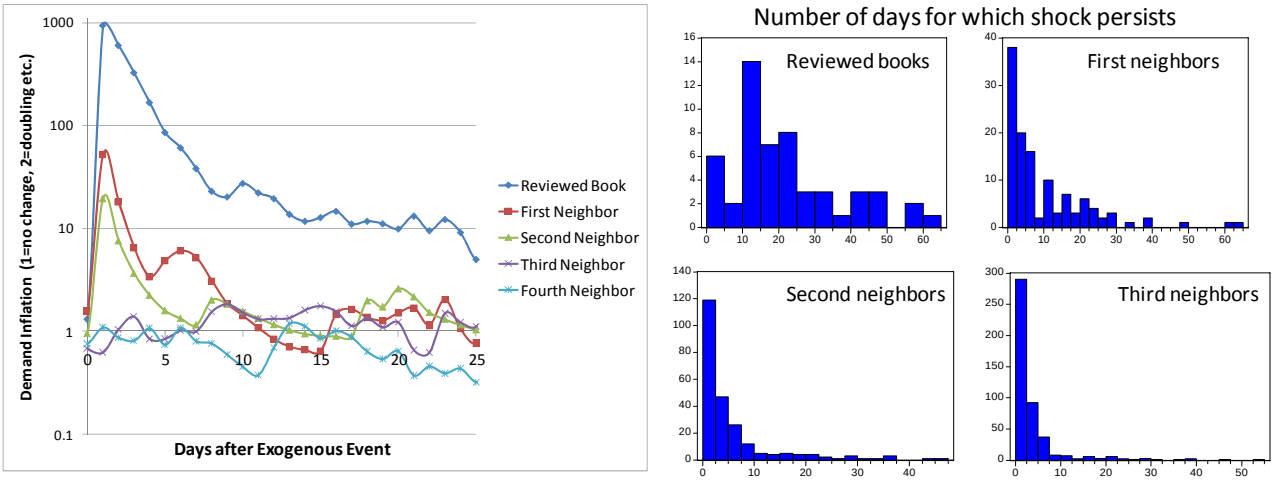

Figure 1: Illustrates the magnitude and persistence of the diffused demand shocks. The graph on the left illustrates the magnitude for a sample set of neighbors relative to their source. The column charts on the right illustrate the number of (y-axis) reviewed books as well as first, second, third neighbors respectively as a function of their persistency duration, or number of post-event days (x-axis) for which demand stayed one standard deviation above pre-event average demand.

Persistence of diffusion: Our evidence suggests that this observed diffused demand shock is remarkably persistent. On average, it takes three weeks for the demand for a reviewed book to return to within one standard deviation of its pre-event levels, although there is considerable variation across books as illustrated by Figure 1. First neighbors experience a persistency of ten days on average, or roughly 50% of the duration of the reviewed book, a clear illustration of the extent to which recommendation networks influence demand over time. Furthermore, 36% of second neighbors have a positive persistency duration, as do *30%* of *third* neighbors and *29%* of *fourth* neighbors. The latter results are striking in their indication of *how far* these demand shocks can travel over the recommendation network.

Other results: In contrasting the changes in demand that were caused by the Oprah events from the New York Times events, we find that while the magnitude of the Oprah events is higher, the New York Times reviews tend to cause demand shifts that were more persistent. Our preliminary results also indicate that the extent of clustering of the recommendation network has a substantial influence on the nature of the diffusion of the demand shock; when the network is highly clustered, the magnitude of diffusion is smaller but the diffusion *persists* for a *longer time*, suggesting that more clustered networks act more like a "fishing net" rather than a "channel for dissemination". We have a variety of other results that we will present in detail if given the opportunity.

Conclusion and Statistical Challenges: Our work presents a new model and evidence of contagion in economic networks, and furthers our understanding of the network characteristics that influence the diffusion patterns of exogenous marketing events. Our key statistical challenges relate to appropriate interpretations of relative inflation in different books' Amazon.com SalesRank, from the fact that a number of our key distributions of shock magnitude and persistence are fat-tailed rather than normal, and in identifying the changes in demand that were on account of the event "flowing" through the network as distinct from those that were caused by changes in the network itself (this is a *co-purchase* network, after all) on account of altered post-event demand patterns. We look forward to the opportunity to present our work and discuss these challenges at the Fifth SCECR.