The Dynamics of Viral Marketing

Presented by:
Group 10

Kevin Huynh
Clayton Schnars
Rahul Mehta
Sahil Pratap

Authors of the paper:
Jure Leskovec - CMU, Lada A. Adamic - University of Michigan, Bernardo A. Huberman- HP Labs
Introduction

Viral Marketing consists of the spreading of information through a network from node to node. In this case, it will be to advertise a series of products.

Word of mouth is a very effective advertising method:

Spreads quickly

People tend to be more trusting of known people (friends, family, etc)

Allows companies to market products indirectly through users
What Will be Studied?

The journal examines first-hand the effect viral marketing has on books, music, and DVDs through a referral system. A discount on the referrer's next purchase is rewarded if the referee purchases the same product.

Components to be discussed:

Effectiveness of word of mouth advertising for hundreds of thousands of products

Characteristics of influences involved in these recommendation networks

Observe the topological behavior of successful vs unsuccessful recommendation graphs
Related Works

Brown and Reingen [1987] determined, after inspecting a referral network for piano lessons, strong ties like family or friendship were extremely effective.

Yang and Allenby [2003] showed that the geographically defined network worked better in regards to the recommendation network of Japanese cars than a demographic network.

Bowman and Narayandas [2001] found that self-reported loyal customers were more likely to talk to others about the products when they were dissatisfied, but, interestingly, they were not more likely to talk to others when they were satisfied.
Epidemic model

Examined by Bailey [1975] and Anderson and May [2002]

Treats graph like an epidemic and nodes are labeled in this order

Susceptible

Infected/Infectious

Immune

Recovered or susceptible, depending on the type of epidemic network

The effectiveness of the epidemic, or how well information spreads in the graph, is determined by an epidemic threshold
The Recommendation Network

- Recommendation referral program by a large retailer
- Books, DVDs, Music, Videos
- First person to click on the recommendation gets 10% off
- The following information is recorded for each recommendation:
  1. Sender Customer ID (shadowed)
  2. Receiver Customer ID (shadowed)
  3. Date of Sending
  4. Purchase flag (buy-bit)
  5. Purchase Date (error-prone due to asynchrony in the servers)
  6. Product identifier
  7. Price
Dataset

- 15,646,121 recommendations
- 3,943,084 distinct users
- 548,523 products were recommended
- Books, DVDs, Music and Videos.

Limitation:

- Recommendation recorded only if the product is bought from the same vendor.
- Recommendation in email form.
Identifying successful recommendations

- A node $i$ first buys a product $p$ at time $t$ and then it recommends it to nodes $j_1, \ldots, j_n$.
- The $j$ nodes can then buy the product and further recommend it.
- Even if all nodes $j$ buy a product, only the edge to the node $j_k$ that first made the purchase will be marked by a **buy-bit** if bought within the first week.
- Identify additional purchases by the presence of outgoing recommendations for a person, since all recommendations must be preceded by a purchase. Referred to as a **buy-edge**.
- Buy edges provide minimum no. of a product bought without discount since there can be customers who do not recommend.
Contd..

- Buy edge: represents a non discount giving edge
- Buy bit: the first person who uses the recommendation within the first week.

Represent the data set as a multigraph:
- The edge \((i, j, p, t)\) indicates that \(i\) recommended product \(p\) to customer \(j\) at time \(t\).
- Note that as there can be multiple recommendations of between the persons (even on the same product) there can be multiple edges between two nodes.
- Edges in a multi graph as recommendations or multi-edges.
- Edge is just to show a connection between two people. Edges between 2 nodes are unique.
The recommendation network

<table>
<thead>
<tr>
<th>Group</th>
<th>( p )</th>
<th>( n )</th>
<th>( r )</th>
<th>( c )</th>
<th>( b_b )</th>
<th>( b_e )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Book</td>
<td>103,161</td>
<td>2,863,977</td>
<td>5,741,611</td>
<td>2,097,809</td>
<td>65,344</td>
<td>17,769</td>
</tr>
<tr>
<td>DVD</td>
<td>19,829</td>
<td>805,285</td>
<td>8,180,393</td>
<td>962,341</td>
<td>17,232</td>
<td>58,189</td>
</tr>
<tr>
<td>Music</td>
<td>393,598</td>
<td>794,148</td>
<td>1,443,847</td>
<td>585,738</td>
<td>7,837</td>
<td>2,739</td>
</tr>
<tr>
<td>Video</td>
<td>26,131</td>
<td>239,583</td>
<td>280,270</td>
<td>160,683</td>
<td>909</td>
<td>467</td>
</tr>
<tr>
<td>Full network</td>
<td>542,719</td>
<td>3,943,084</td>
<td>15,646,121</td>
<td>3,153,676</td>
<td>91,322</td>
<td>79,164</td>
</tr>
</tbody>
</table>

Table 1: Product group recommendation statistics. \( p \): number of products, \( n \): number of nodes, \( r \): number of recommendations, \( c \): number of edges, \( b_b \): number of buy bits, \( b_e \): number of buy edges.

- Small no. of DVD options but account for majority of recommendations
- DVDs: 10 recommendations per node
- Books and Music: 2 recommendations per node
- Video: 1 recommendation per node.
- Music nodes reached the same number of users as DVDs but took 1/5th fewer recommendations.
- Based on the small number of unique edges compared to no. of nodes, the networks are highly disconnected.
The Largest Weakly Connected Component

- The largest weakly connected component is really small compared to the actual sample.
- the LCC contains the 2.54% of nodes contribute to 52.9% of the recommendation.

<table>
<thead>
<tr>
<th>Group</th>
<th>$n_c$</th>
<th>$r_c$</th>
<th>$e_c$</th>
<th>$b_{bc}$</th>
<th>$b_{ec}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Book</td>
<td>53,681</td>
<td>933,988</td>
<td>184,188</td>
<td>1,919</td>
<td>1,921</td>
</tr>
<tr>
<td>DVD</td>
<td>39,699</td>
<td>6,903,087</td>
<td>442,747</td>
<td>6,199</td>
<td>41,744</td>
</tr>
<tr>
<td>Music</td>
<td>22,044</td>
<td>295,543</td>
<td>82,844</td>
<td>348</td>
<td>456</td>
</tr>
<tr>
<td>Video</td>
<td>4,964</td>
<td>23,555</td>
<td>15,331</td>
<td>2</td>
<td>74</td>
</tr>
<tr>
<td>Full network</td>
<td>100,460</td>
<td>8,283,753</td>
<td>521,803</td>
<td>8,468</td>
<td>44,195</td>
</tr>
</tbody>
</table>

Table 2: Statistics for the largest connected component of each product group. $n_c$: number of nodes in largest connected component, $r_c$: number recommendations in the component, $e_c$: number of edges in the component, $b_{bc}$: number of buy bits, $b_{ec}$: number of buy edges in the largest connected component, and $b_{bc}$ and $b_{ec}$ are the number of purchase through a buy-bit and a buy-edge, respectively.
Recommendation network over time

Small World networks

Shows growth of LCC is quadratic

Inset shows linear time increase in nodes ~165,000 which ends up as evidence of a non viral spread
Growth of the LCC

(a) LCC growth
(b) Sender in LCC
(c) Sender outside LCC

a. The distribution of sizes of components when they are merged into the largest connected component.
b. Same as (a), except limited to recommendations sent from LCC to an external Node
c. Only the nodes whose recommendation senders were outside the LCC.

X-axis: No. of nodes.
Y-axis: number of components of size N that were merged over time with the largest component.
Propagation of recommendations

- Forwarding recommendations
- Not everyone forwards after accepting recommendation
- About only a third recommended it further.
- Forward recommendations really high for DVDs and low for books.

- Cumulative out-degree distribution
- Exponential drop-off at around 100 recommendations.
- Deeper an individual is in the cascade, if they choose to make recommendations, they tend to recommend to a greater number of people on average.
- Probability of forwarding a recommendation is low declines after an initial increase as one gets deeper into the cascade
Identifying Cascades

- Each cascade is a network consisting of customers (nodes) who purchased the same product as a result of each other’s recommendations (edges).
- Customers continue forwarding recommendations, they contribute to the formation of cascades.
- To identify cascades, we track successful recommendations as they influence purchases and further recommendations.
- Recommendation to be successful if it reached a node before its first purchase.
- In case of multiple recommendations, only the first one is considered.
- This way we make the network time increasing or causal i.e. for each node all incoming edges come before all outgoing.
Observations

• Most product recommendation networks consist of a large number of small disconnected components where we do not observe cascades.
• Then there is usually a small number of relatively small components with recommendations successfully propagating.
• The most active customer made 83,729 recommendations and purchased 4,416 different items.
• The heavy tailed distribution of cascade sizes
Shows size distribution of cascades (size of cascade vs. count). Bold line presents a power-fit.
Exchanged recommendations between pairs:

- 39% of pairs of people exchanged just a single recommendation.
- This number decreases for DVDs to 37%, and increases for books to 45%.
- Notice that one gets much stronger decay exponent (distribution has weaker tail) of -2.7 for books and a very shallow power-law exponent of -1.5 for DVDs.
- This means that even a pair of people exchanged more DVD than book recommendations.
Probability of Buying versus Number of Incoming Recommendations

• Saturation Point
  • How to determine the maximum limit for the number of recommendations to be sent before they lose their value

• How to Determine the probability
  • $n$ - the number of observations
  • $m$ - the number of successes
  • $k (= n - m)$ - the number of failures
  • estimated probability of purchasing is $p = m/n$

• The values vary from product to product
• Books - Peak in probability of buying at 2 incoming recommendations and then a slow drop

• DVD’s – Saturation at about 10 DVDs
Success of Subsequent Recommendations

- Trusted Recommender vs Spammer

(a) Books

(b) DVD
Success of Outgoing Recommendations

• Sender’s Viewpoint
  • Probability of getting a 10% credit?
  • Influence of outgoing recommendations on purchases?
  • To be selective or spam?

• Results
  • For books, music, and videos, the number of purchases soon saturates while DVDs exhibit different behavior, with the expected number of purchases increasing throughout.
  • books, videos, and music have qualitatively similar trends as earlier and the success of DVD recommendations saturates as well.
The difference in the curves for DVD recommendations points to the presence of **collisions** in the dense DVD network, which has 10 recommendations per node and around 400 per product, which is an order of magnitude more than other product groups.
Probability of Buying Given the Total Number of Incoming Recommendations

• What is level of involvement of people in the recommendation network?
  • Most people are not expected to be highly involved
  • Some extreme cases will purchase all recommended products

• Ways to determine the level of involvement
  • The probability of buying as a function of the number of different products recommended
  • The probability of buying as a function of the total number of incoming recommendations

• Results
  • For books and music the probability of buying is the highest when a person got recommendations on just 1 item
  • In case of movies (DVDs and videos), with more recommendations the person is more likely to buy.
The probability of buying as a function of the number of different products recommended

The probability of buying as a function of the total number of incoming recommendations
Timing of Recommendations and Purchases

• The recommendation referral program encourages people to purchase as soon as possible after they get a recommendation since this maximizes the probability of getting a discount.

• *thinking time – what does it mean?*
• Variation in intensity by time of day for three different activities in the recommendation system: recommendations, all purchases, and finally just the purchases which resulted in a discount.

![Graphs showing recommendations, all purchases, and purchases with discount over time of day.](image)

- The recommendations and purchases follow the same pattern. The number of purchases with discount is the highest when the number of purchases is small.
Recommendations and Communities of Interest

- Find communities using Q Modularity
- Identify interests using formula: \[ x_g \times p_c \pm \sqrt{x_g \times p_c \times (1 - p_c)} \]
  - \( x_g \) = Number of recommendations sent in community \( g \)
  - \( p_c \) = Proportion of recommendations belonging to category \( c \)
- when \( p_c \) is large \( c \) is an interest of \( g \)
<table>
<thead>
<tr>
<th># nodes</th>
<th># senders</th>
<th>topics</th>
</tr>
</thead>
<tbody>
<tr>
<td>735</td>
<td>74</td>
<td>books: American literature, poetry</td>
</tr>
<tr>
<td>710</td>
<td>179</td>
<td>sci-fi books, TV series DVDs, alternative rock music</td>
</tr>
<tr>
<td>667</td>
<td>181</td>
<td>music: dance, indie</td>
</tr>
<tr>
<td>653</td>
<td>121</td>
<td>discounted DVDs</td>
</tr>
<tr>
<td>541</td>
<td>112</td>
<td>books: art &amp; photography, web development, graphical design, sci-fi</td>
</tr>
<tr>
<td>502</td>
<td>104</td>
<td>books: sci-fi and other</td>
</tr>
<tr>
<td>388</td>
<td>77</td>
<td>books: Christianity and Catholicism</td>
</tr>
<tr>
<td>309</td>
<td>81</td>
<td>books: business and investing, computers, Harry Potter</td>
</tr>
<tr>
<td>192</td>
<td>30</td>
<td>books: parenting, women’s health, pregnancy</td>
</tr>
<tr>
<td>163</td>
<td>48</td>
<td>books: comparative religion, Egypt’s history, new age, role playing games</td>
</tr>
</tbody>
</table>
Table VI. Statistics by Book Category

- $n_p$: number of products in category,
- $n$: number of customers,
- $cc$: percentage of customers in the largest connected component,
- $r_{p1}$: avg. # reviews in 2001–2003,
- $r_{p2}$: avg. # reviews 1st 6 months 2005,
- $v_{av}$: average star rating,
- $c_{av}$: average number of people recommending product,
- $c_{av}/r_{p1}$: ratio of recommenders to reviewers,
- $p_m$: median price,
- $b$: ratio of the number of purchases resulting from a recommendation to the number of recommenders.

* denotes statistical significance at the 0.05 level, ** at the 0.01 level.

<table>
<thead>
<tr>
<th>Category</th>
<th>$n_p$</th>
<th>$n$</th>
<th>$cc$</th>
<th>$r_{p1}$</th>
<th>$v_{av}$</th>
<th>$c_{av}/r_{p1}$</th>
<th>$p_m$</th>
<th>$b \times 100$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Books general</td>
<td>370230</td>
<td>2,860,714</td>
<td>1.87</td>
<td>5.28</td>
<td>4.32</td>
<td>1.41</td>
<td>14.95</td>
<td>3.12</td>
</tr>
<tr>
<td>Literature</td>
<td>41,682</td>
<td>502,179</td>
<td>3.06</td>
<td>13.09</td>
<td>4.30</td>
<td>0.57</td>
<td>11.87</td>
<td>2.82*</td>
</tr>
<tr>
<td>Romance</td>
<td>6,317</td>
<td>60,902</td>
<td>5.65</td>
<td>12.81</td>
<td>4.17</td>
<td>0.52</td>
<td>6.99</td>
<td>1.78**</td>
</tr>
<tr>
<td>Entertainment</td>
<td>18,724</td>
<td>258,142</td>
<td>3.65</td>
<td>3.48</td>
<td>4.29</td>
<td>2.26</td>
<td>13.97</td>
<td>2.66*</td>
</tr>
<tr>
<td>Health/Body</td>
<td>33,751</td>
<td>572,704</td>
<td>1.54</td>
<td>4.34</td>
<td>4.41</td>
<td>2.39</td>
<td>13.96</td>
<td>3.04</td>
</tr>
<tr>
<td>History</td>
<td>28,458</td>
<td>28,3406</td>
<td>2.74</td>
<td>4.34</td>
<td>4.30</td>
<td>1.27</td>
<td>18.00</td>
<td>2.84</td>
</tr>
<tr>
<td>Home/Garden</td>
<td>19,024</td>
<td>180,009</td>
<td>2.91</td>
<td>1.78</td>
<td>4.31</td>
<td>3.48</td>
<td>15.37</td>
<td>2.26**</td>
</tr>
<tr>
<td>Entertainment</td>
<td>18,724</td>
<td>258,142</td>
<td>3.65</td>
<td>3.48</td>
<td>4.29</td>
<td>2.26</td>
<td>13.97</td>
<td>2.66*</td>
</tr>
<tr>
<td>Medicine</td>
<td>16,047</td>
<td>175,520</td>
<td>1.08</td>
<td>1.41</td>
<td>4.40</td>
<td>4.19</td>
<td>39.95</td>
<td>5.68**</td>
</tr>
<tr>
<td>Engineering</td>
<td>10,312</td>
<td>107,255</td>
<td>1.30</td>
<td>1.43</td>
<td>4.14</td>
<td>3.85</td>
<td>59.95</td>
<td>4.10**</td>
</tr>
<tr>
<td>Law</td>
<td>5,176</td>
<td>53,182</td>
<td>2.64</td>
<td>1.89</td>
<td>4.25</td>
<td>2.67</td>
<td>24.95</td>
<td>3.66*</td>
</tr>
<tr>
<td>Nonfiction</td>
<td>55,868</td>
<td>560,552</td>
<td>2.03</td>
<td>3.13</td>
<td>4.29</td>
<td>1.89</td>
<td>18.95</td>
<td>3.28**</td>
</tr>
<tr>
<td>Reference</td>
<td>26,834</td>
<td>371,959</td>
<td>1.94</td>
<td>2.49</td>
<td>4.19</td>
<td>3.04</td>
<td>17.47</td>
<td>3.21</td>
</tr>
<tr>
<td>Biographies</td>
<td>18,233</td>
<td>277,356</td>
<td>2.80</td>
<td>7.65</td>
<td>4.34</td>
<td>0.90</td>
<td>14.00</td>
<td>2.96</td>
</tr>
</tbody>
</table>
Products and Recommendations

- Long tail: top 1000 products were 27% of sales
  - 67% had a single purchase = 30% of sales
Modeling Recommendation Success

\[ s = \exp \left( \sum_i \beta_i \log(x_i) + \epsilon_i \right) \]

\[ X_i \text{ product attributes:} \]
- \( n \) unique nodes
- \( n_s \) senders
- \( n_r \) receivers
- \( r \) recommendations
- \( e \) edges
- \( p \) price
- \( v \) reviews
- \( t \) average rating

<table>
<thead>
<tr>
<th>Variable</th>
<th>Books Coefficient ( \beta_i )</th>
</tr>
</thead>
<tbody>
<tr>
<td>const</td>
<td>1.317 (0.0038)**</td>
</tr>
<tr>
<td>( n )</td>
<td>-0.579 (0.0060)**</td>
</tr>
<tr>
<td>( n_s )</td>
<td>0.144 (0.0018)**</td>
</tr>
<tr>
<td>( n_r )</td>
<td>-0.006 (0.0064)</td>
</tr>
<tr>
<td>( r )</td>
<td>0.062 (0.0084)**</td>
</tr>
<tr>
<td>( e )</td>
<td>0.383 (0.0106)**</td>
</tr>
<tr>
<td>( p )</td>
<td>0.013 (0.0003)**</td>
</tr>
<tr>
<td>( v )</td>
<td>-0.003 (0.0001)**</td>
</tr>
<tr>
<td>( t )</td>
<td>-0.001 (0.0006)*</td>
</tr>
</tbody>
</table>
Fig. 17. A Bayesian network showing the dependencies between the variables.
Conclusion

- SIRS model assumption contradictions
- Cascading and communities
- Recommendation success
- Not so viral marketing