How To Make Your Marketing Campaign Viral

Simone Maggi

ABSTRACT

Nowadays, social networks have become a useful tool for marketing strategist to achieve great public visibility. Thanks to their appearance viral marketing strategies start to gain prestige. The viral marketing dream is to popularize the brands as broadly as possible leveraging on the interest of individuals in sharing content to his restricted circle of friends. A successful viral marketing campaign surely depends on the intrinsic value of the brands or the creativity in promoting it, but partially it depends also on the underlying network of friendships exploited by this kind of marketing. This article will focus on the importance of the network structure in promoting successfully a brand.

Keywords: Innovation Cascade, Complex Network

INTRODUCTION

The evolution of a viral marketing campaign will be examined in a simplified environment. A market with only one competitor besides our brand is chosen. Moreover, individuals will choose our brand or the competitor's one considering only the choices of their friends. The underlying network of friendship used in the analysis is a small subset of Deezer, composed of approximately thirty thousand of individuals interconnected.

GENERIC DESCRIPTION OF THE NETWORK

The network dataset is free to download at http://snap.stanford.edu/data and consists of 28281 Deezer's users connected by 92752 edges. Since it is a social network, we expect to see the common proprieties that characterize this kind of networks.

Small World

First of all we check that our network has the small world property. The global clustering coefficients is $C_g = 0.096$ much more greater than the density $\rho = 0.00023$. The plot 1 shows the distribution of the shortest path. The red line is the average path length $< g >= 6.45 \simeq \frac{\log(n)}{\log(\frac{2m}{n})} = 5.44$. Hence we conclude that our network is a small world.

Degree Distribution

The network contains hubs, hence we expect a power law distribution of degrees. Figure 2 shows the log-log plot of the degree distribution. A linear model fits reasonably the points, confirming a power law distribution with density $f_x(x) = 11.86x^{-2.7}$. The plot 3 shows the fit in linear-linear scale.

Degrees Correlations

The degree correlation function is plotted in log-log coordinates in figure 4. The red horizontal line at $\frac{<k^2>}{<k>} = 15.91$ represents the ideal behaviour of a neutral network,i.e. a network without degree correlation. A linear fit,the green line, on the degree correlation function shows that our network is slightly assortative. Nevertheless the slope coefficient of the fit has a p-value of $p = 0.02$, greater than a significance level $\alpha = 0.01$. This suggests that our fit could depends only on a constant coefficient becoming more similar to the red horizontal line.

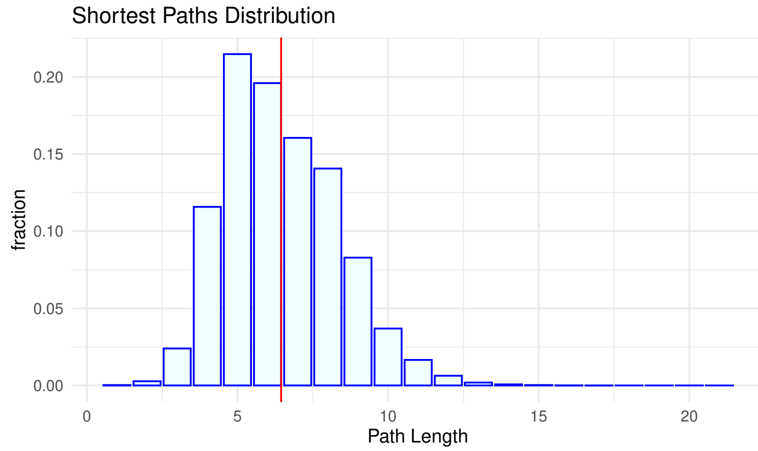

Figure 1. shortest path distribution

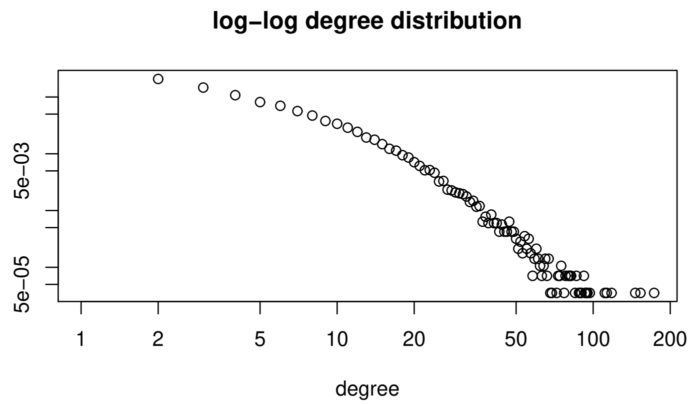

Figure 2. Degree distribution in log-log scale.

Rich Club Effect

The function plotted in 5 describes the density of the sub-graph induced by the fraction r of the most popular vertices. The decreasing behaviour shows that the popular nodes are more interconnected than the other nodes of the network: we are facing a rich club effect.

Homophily

Jointly to the network structure data, it has been provided a set of features for each vertex. Among them, there is the sex of each user. With this information, we can investigate if in this network there is the presence of homophily between the sex groups. For this aim, we divide the users by sex, and we test if the edges between the two groups are significantly less than a randomly generated network with the same number of nodes and edges. The fraction of between-groups edges in the random network is $u_0 = 2p(1-p) = 0.4935$, meanwhile the actual fraction of such edges in our network is $u = 0.4749 < u_0$, where p is the proportion of males. To ensure that it is smaller enough a statistical test must be performed. Call X_{ij} a random variable associated to each edge which is 1 if the edge (i, j) is between the first group G_1 and the second group G_2, and zero otherwise. If the network does not present homophily, the $EX_{ij} = 2p(1-p) = u_0$, i.e. the expected fraction of between group edges is equal to the randomly generated network one. Hence we test $H_0 : EX_{ij} = u_0$ against $H_1 : EX_{ij} < u_0$, where an unbiased estimator for EX_{ij} is $u = \frac{1}{m}\sum_{i,j\in E(G)} X_{ij}$, the empirical mean of X_{ij} obtained in our network. Under H_0 it is true that $y = \sqrt{m}(u-u_0) \xrightarrow{m\hookrightarrow\infty} Z$, with $Z \sim N(0,(u_0(1-u_0))^2)$. Hence we find θ_0 such that $P(rejectH_0|H_0true) = P(u < \theta_0|H_0true) = \alpha$, with the significance level $\alpha = 0.05$. But

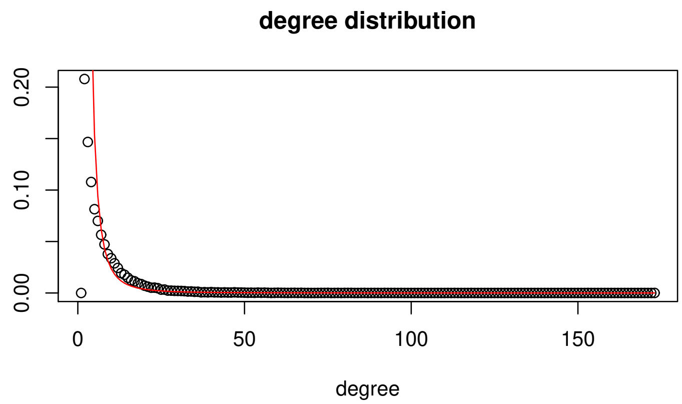

Figure 3. Degree distribution with fitted density.

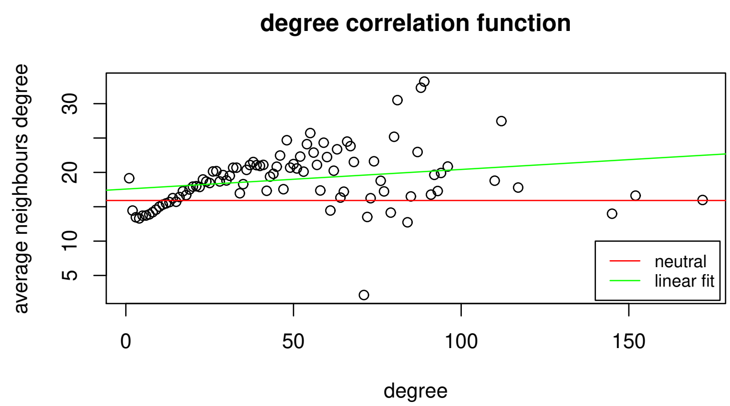

Figure 4. Degree correlation function.

$P(u < \theta_0 | H_0 true) = P(\sqrt{m}(u - u_0) < \sqrt{m}(\theta_0 - u_0) | H_0 true) \simeq P(Z < \sqrt{m}(\theta_0 - u_0) | H_0 true)$. Assuming that our approximation is good, we find $\theta_0 = 0.4903 > 0.4749 = u$. Concluding, since u is smaller than the threshold θ_0 we can reject H_0 and accept the hypothesis of homophily.

SIMULATION OF THE MARKETING CAMPAIGN

In order to study the impact of the underlying network structure in the spreading of the campaign, we suppose that individuals choose to adopt our brand if the fraction of their friends that are adopting already it is bigger than a threshold q. The threshold depends on the advantage of choosing our brand with respect to the competitive one. The simulation is performed as follow:

1. Let S be the initial nodes that already adopted our brand. Call C_t a vector whose i-th element is 1 if node i is adopting our brand at time t, 0 if not. Define A to be the adjacency matrix of our graph and D^{-1} to be a $n \times n$ diagonal matrix containing the inverse of the degree of each node.
2. Compute for each i-th element of C_{t+1} the value $p_i = \sum_{j=0}^{n} a_{ij} C_t^{(j)}$, and set $C_{t+1}^{(i)} = 1$ if $p_i > q$, or zero otherwise.
3. If the number of individuals adopting our brand is increased with respect to the previous step, repeat 2, otherwise end the simulation.

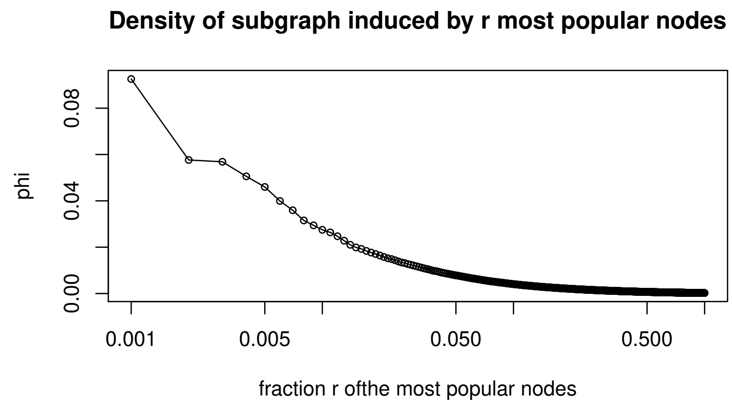

Figure 5. plot of $\phi(r)$.

How To Be Viral

It's clear that in this very simple model we can improve our impact on the market by decreasing the threshold q, i. e. making our brand a lot better than the competitors, but this is not always possible. Otherwise we can convince a specific set of initial nodes to use our brand at the start, in order to maximize the diffusion of our marketing campaign.
Imagine that we are introducing our brand in a market already dominated by the competitor, but it is true that our products have a slightly better quality such that the threshold $q = \frac{1}{3}$. Suppose that we have kept apart some money for the marketing campaign and we are willing to pay a initial set of influencers who could be able to viralize our marketing campaign: what influencers should we choose?
Plot 6 compare the evolution of the spreading of our brand depending on the type of starting nodes chosen. For each simulation the number of starting nodes is fixed to $k = 200$. The green line shows the evolution of the simulation choosing the k most populat nodes, the red is obtained choosing the k most close to all the other nodes, and the blue is obtained choosing k random nodes. The degree centrality has outperformed the closeness centrality reaching 22895 versus only 1331 users. Although, both degree and closeness centralities are more effective than choosing k random nodes. The explanation is obvious and reflects what happen frequently in real marketing campaign: the marketing strategists always include in their campaigns VIP and popular faces who are in touch with big part of the public.

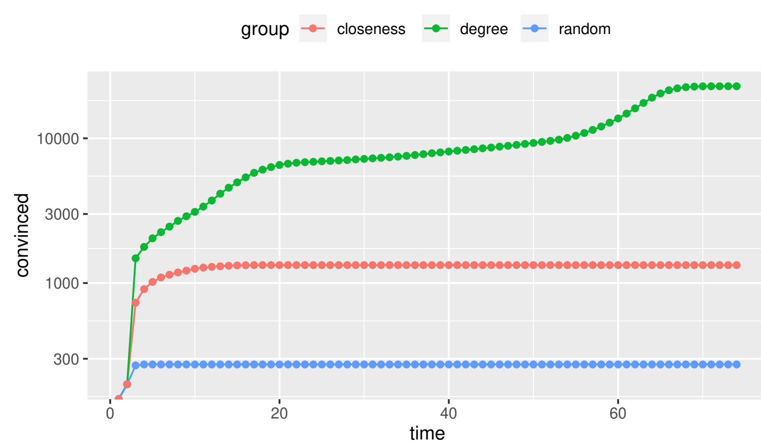

Figure 6. Comparison of the different simulation. Choosing the k nodes with higher degrees outperformed the other choices.

k-core decomposition

In the last plot 7 it is used a different approach: as starting nodes are chosen the 10-core of the graph, composed by 264 nodes. The k-core of a graph is the maximal sub-graph in which each vertex has at least degree k. This idea works very well when the graph exhibit a good assortativity, hence a core-perifery structure. Indeed, using as starting nodes the ones in the core of the graph that are densely interconnected, the cascade should quickly infect the whole network. Unfortunately the network under exam is very slightly assortative, as we have seen before, hence this method does not work better than the others. Nevertheless it gives good cascade performances reaching 4946 individuals.

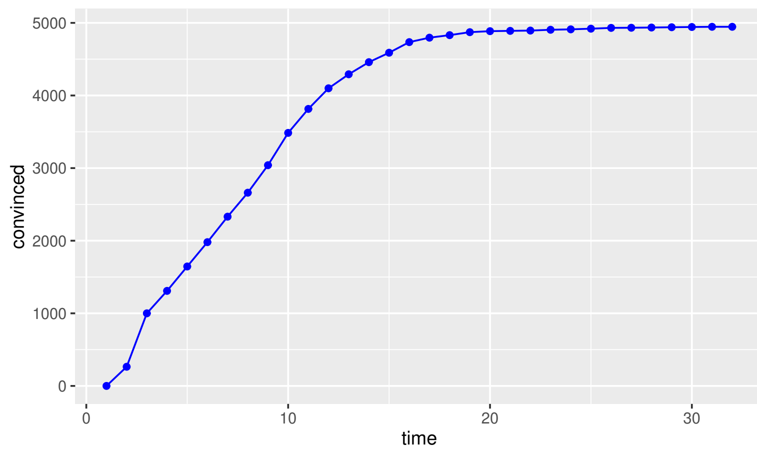

Figure 7. 10-core of the graph as starting nodes.

VISUALIZATION

Concluding, in 8 can be observed a visualization of the graph. The structure appears to be quite polarized among the three biggest communities of the graph: the purple contains about 6000 nodes, the other two togheter contain about 12000 nodes. The communities are obtained with the modularity algorithm ("Vincent D Blondel).

REFERENCES

["Vincent D Blondel] "Vincent D Blondel, Jean-Loup Guillaume, R. L. E. L. "fast unfolding of communities in large networks". *"in Journal of Statistical Mechanics: Theory and Experiment"*.

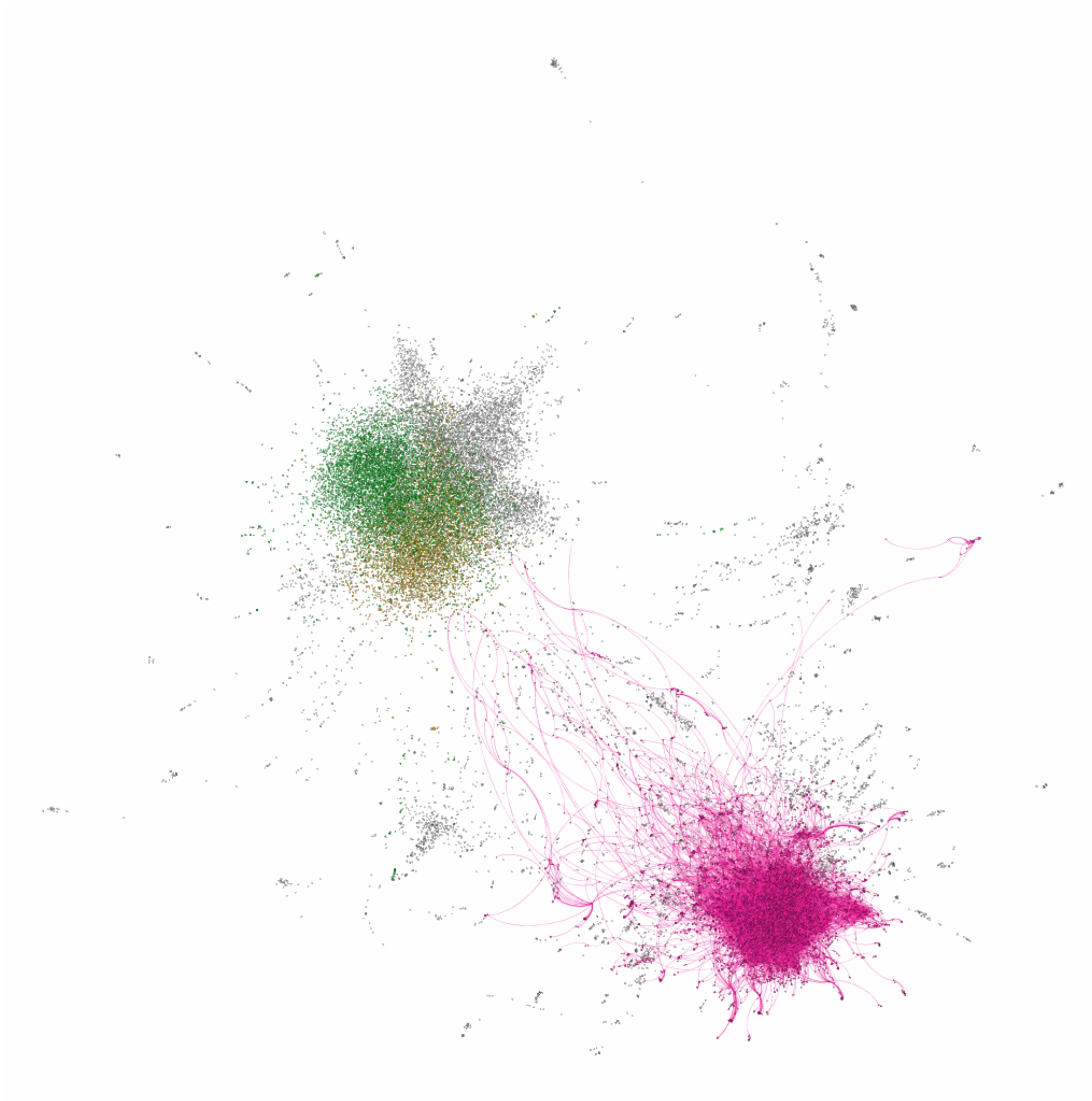

Figure 8. Graph visualization. Each color is a community.