Understanding the Connections between Networks and Choice Modeling

Xuling Yang

Mentor: Aparajithan Venkateswaran

Department of Statistics,

University of Washington, Seattle, Washington 98105, the United States

March 29, 2021

This paper introduces concepts and theories that I learned in the SPA DRP program, explains them in real-life examples, and also demonstrates my own ideas and thoughts of understanding the connections between networks and choice modeling.

1 Introduction

Choice is something that we will always encounter. In fact, our whole life is formed by countless decision-making moments: from trivial matters such as which shirt to wear or which pizza to order, to life-changing crossroads such as which career to take or even which person to marry to. It seems so necessary for us to be able to pick the "right" option which will bring the greatest benefit. In

choice models, this kind of benefit is defined as "utility".

However, in most of the cases, it is difficult to tell whether an option is more beneficial than another before the choice is actually made, especially when the difference between options are not that large. Thus, here comes a genius idea – why don't we build a model to work out and compare the exact numerical value of the total utility of the system when each option is chosen, so that the abstract decision-making is turned into a solid math problem? Thus, we come up with the idea of "Discrete Choice Models".

2 Discrete Choice Models

Discrete Choice Models, short for DCM, is a family of mathematical models which compare the utility of each option in the system and figure out the probability that whether the option will be chosen or not. By using the choice model, we are able to "obtain market-level predictions" by analyzing the preference of specific customers, or "give forecasts over time" by gaining information from previous data.

In this research, our exploration of Discrete Choice Models depends on the *"Discrete Choice Methods with Simulation"* written by Kenneth Train. [1]

2.1 Properties

Here are some factors and properties that we need to know before building the model.

For any decision-making event, there are always two subjects – decision maker and choice set. A decision maker can be you and me, or anyone who is going to make a choice, and the choice set is

a set of alternatives that are possible to be chosen. The choice set in the DCM should follow three characteristics: mutually exclusive, exhaustive, and finite. Mutual exclusiveness means that the decision-maker can pick only one option from the choice set; exhaustiveness makes sure that any possible alternative is included in the choice set so that the decision maker is necessary to choose one option; to be finite means that there should be an upper limit of the number of alternatives in the set, and that the set should be countable. The first two criteria are not restrictive—to achieve different research goals, different definitions can be given to the choice set to meet the requirements. However, "finite" is restrictive and should never be changed, since DCM does not share the same rules and results with the continuous choice models or those with infinite options—those are not "discrete" at all.

To understand these criteria and restrictions, we can take a look at the example of two apples: one is red, and the other is green. The question is, "between the two apples, which one do you like better?" To meet exclusiveness and exhaustiveness, the alternatives should be "red only" and "green only". However, when the apples are put in the supermarket, and people have the right to choose whatever product they want, then the alternatives should be defined as "red only", "green only", "both red and green", and "neither of them". In this case, the situation is becoming more complicated since it brings more questions like "how much people love apple more than other fruits" and relative data should be available for further analysis. Therefore, the definition of the choice set can be changeable and flexible to meet the first two criteria for different goals, though most of the time there exists a tradeoff between these definitions. However, these are all based on the "finite" characteristic, for which we have to count apples by a whole, or a combination, instead of cutting them into pieces of countless different shapes and make each of them an option.

2.2 Utilities and Probabilities

Utility, as we have just mentioned, is the usefulness of an alternative, or the benefit the alternative will bring if it is chosen. The utility of an alternative is decided by two kinds of factors: the observed factors, denoted by the vector x; and the unobserved factors, whose effects all represented by the variable ε.

Therefore, the utility function can be written as

$$U = \beta' \mathbf{x} + \varepsilon \tag{1}$$

, where represents the vector of parameters.

By using the utility function, we can then write the probability of the alternative being chosen as

$$\begin{aligned} P &= \int I[\beta' \mathbf{x} + \varepsilon > 0] f(\varepsilon) d\varepsilon \\ &= \int I[\varepsilon > -\beta' \mathbf{x}] f(\varepsilon) d\varepsilon \end{aligned} \tag{2}$$

, where ε is assumed as distributed logistic, whose density written as $f(\varepsilon) = \frac{e^{-\beta}}{(1+e^{-\beta})^2}$ with cumulative distribution $F(\varepsilon) = \frac{1}{1+e^{-\beta}}$. The I here is an indicator function. When the statement in the square is true, $I = 1$, otherwise $I = 0$. This shows that only when the utility is positive does it

contributes to the probability of being chosen.

$$\begin{aligned} P &= \int_{\varepsilon=-\beta'\mathbf{x}}^{\infty} f(\varepsilon)d\varepsilon \\ &= 1 - F(-\beta'\mathbf{x}) \\ &= 1 - \frac{1}{1+e^{\beta'\mathbf{x}}} \\ &= \frac{e^{\beta'\mathbf{x}}}{1+e^{\beta'\mathbf{x}}} \end{aligned} \tag{3}$$

, which turns out to be a logistic function.

This is the probability distribution for choosing one single item. However, when there are multiple items in the same system, we need to conduct pairwise comparison on the utilities to obtain the probabilities. In the common sense, people always would like to pick the option with greater utility. The greater the utility is, the more likely it will be chosen. Therefore, we define

$$\begin{aligned} P(i \to j) &= Prob(U_i > U_j) \\ &= Prob(U_i - U_j > 0) \end{aligned} \tag{4}$$

as the probability that option i beats j.

It also indicates that when doing pairwise comparison, "only differences in utility matter" and that "the scale of utility is arbitrary". These two statements will also be reminded later in section 4 at page 11

In fact, in reality, not all people prefer to choose the option with the greatest utility. To maximize the utility is only one of Discrete Choice Method. These models are called "Random Utility Models". Bradley-Terry-Luce (BTL) model is one of the RUMs, where the probability that the

decision-maker chooses i over j is given by

$$\begin{aligned} P(i \to j) &= \frac{e^{U_i}}{e^{U_i} + e^{U_j}} \\ &= \frac{1}{1 + e^{-(U_i - U_j)}} \end{aligned} \tag{5}$$

, which is obviously a logistic distribution. We will also come back to this in section 4. (equation 8)

3 Networks and Rankings

A network, by definition, is "a collection of vertices joined by edges". [Network basics] A network can also be represented by a graph G = (V, E), where V is the set of vertices, and E is the set of edges. There are lots of different kinds of networks or graphs, such as simple graphs and multigraphs; undirected graphs and directed graphs; cyclic graph or acyclic graph. Typically, a network can be represented in three ways: the adjacency matrix, the adjacency list, and the edge list. There are also some structural measures of networks, such as degrees, geodesic paths, and clustering coefficients.

However, despite all these basic ideas of networks, what we truly pay attention to is the strong function of the networks as a tool of ranking. It is true that discrete choice models are all what we need to do pairwise comparisons of utilities, but when there exists a great number of options in the choice set, it would be much more complicated and heavy work to work out and rank all the choice probabilities. Thus, it would be perfect if we can find some connections between choice modeling and networks—the natural ranking tool. In this chapter, we are going to talk about how rankings of vertices in a network are measured, as well as three typical ranking methods.

In this research, our exploration of Networks and Rankings depends on the *"Network Analysis and Modeling, CSCI 5352 Lecture Notes"* by Clauset and Larremore. [2]

3.1 Measurement of Rankings: Structural Importance

For rankings in networks, the score of each vertex actually represents its importance level in this system. And the importance level itself, in a static network with no dynamical process, depends mostly on the position of the vertex in the network structure. This is what we called "Structural Importance".

Typically, the structural importance is defined by the centrality level of the vertex in the network. The more central the vertex is, the more important it is. It is not hard to understand if we think it as a real-life network, where each vertex represents a person, and each edge represents the relationship between people. It would be not surprising to find that in each network, there is someone who knows a lot of people, and thus gathers everyone together as the core of the group, while there is also someone has few links with other members in the group, and thus is relatively isolated near the margin of the network. Therefore, the cores, with higher centrality, have greater importance than the isolates in network structure. Moreover, we can also find that the centrality level has something to do with the relationships a person has. The stronger the relationships are, and the higher the quantity is, the more important the person is. Therefore, in the next section, we focus mostly on the weight and degree to work out the function of ranking score.

3.2 Three Ranking Methods by Structural Importance

In this section, we are introducing three typical ranking methods depending on structural importance: degree centrality, eigenvector centrality, and PageRank.

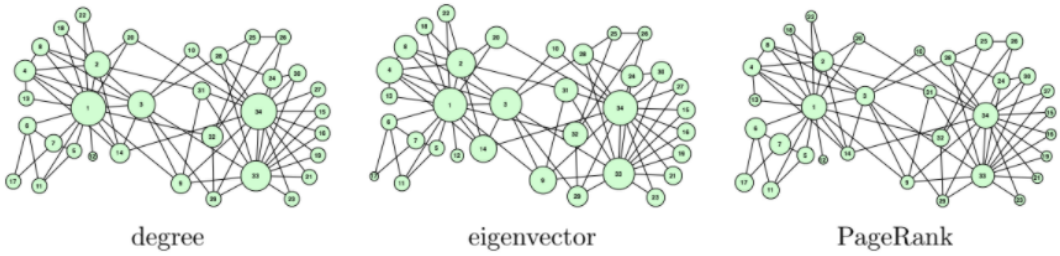

Figure 1: (a) *degree centrality:* k/m (b) *eigenvector centrality:* $x_i^{(t+1)} = \sum_{j=1}^{n} A_{ij} x_j^{(t)}$ (c) *PageRank:* $x_i = \alpha \sum_{j=1}^{n} A_{ij} \frac{x_j}{k_j^{out}} + \beta$ [2]

The degree centrality is the simplest one of the three ranking methods. It depends only on the proportion of the degree a vertex has out of the total degrees. It can only be used for undirected graphs.

The eigenvector centrality uses the recursive method to get the final score of each vertex. At first, each vertex is given a same beginning score, which is usual very small, like "1". Then the iterations begin and each vertex keeps giving each of its neighbors the value of its score times the weight of the edge. In each iteration, each vertex sums up their score from last iteration and the scores given by their neighbors. And when the score of each vertex converges to a fixed number, we get the final ranking scores.

The PageRank is something that we will definitely encounter when we are using search engines, like Google. In this system, we can view each net page as a node and each hyperlink on the page as an edge. When we click on the hyperlink, we trigger the edge and it will direct us to another page. Since this is a one-way direction, it is not hard to understand that the PageRank applies to directed graphs. Its algorithm is a bit similar to that of the eigenvector centrality. However, since the edges are directed in this case, each vertex gives its original scores away through its out-going edges by equal proportion, and receives new scores from its in-coming edges.

The following scatter plots show that there is positive correlation between any two of the three

ranking methods. From the graphs (figure 1), plots (figure 2), and table (figure 3), we can see that these ranking methods have similar effects and results. [2]

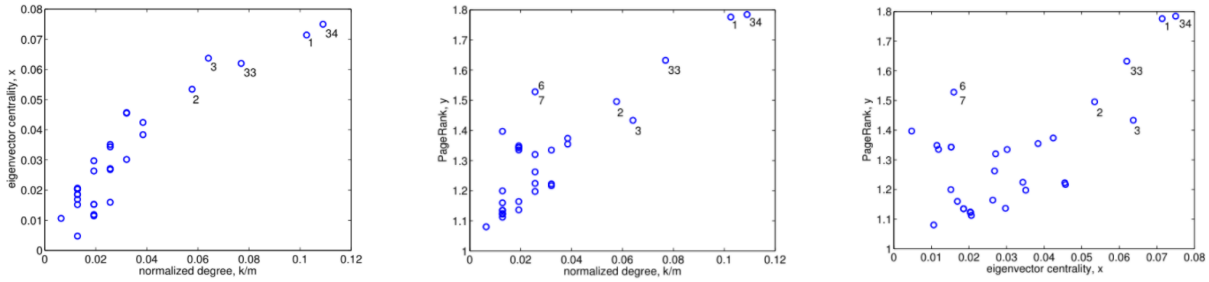

Figure 2: Scatter Plots

	degree k/m		eigenvector		PageRank	
1^{st} largest	34	(0.1090)	34	(0.0750)	34	(1.7843)
2^{nd} largest	1	(0.1026)	1	(0.0714)	1	(1.7758)
3^{rd} largest	33	(0.0769)	3	(0.0637)	33	(1.6324)
4^{th} largest	3	(0.0641)	33	(0.0620)	6	(1.5280)
5^{th} largest	2	(0.0577)	2	(0.0534)	7	(1.5280)

Figure 3: Table

However, since these ranking methods depends mostly on degrees and geodesic distances, we can get only very coarse ordinary rankings, but the score itself has no actual meaning, and thus cannot represent the utility of an option. It seems that it does not make any sense if we use these networks with rankings to represent the distribution of utility models. Then how can we come to the reconciliation?

Fortunately, we find the SpringRank.

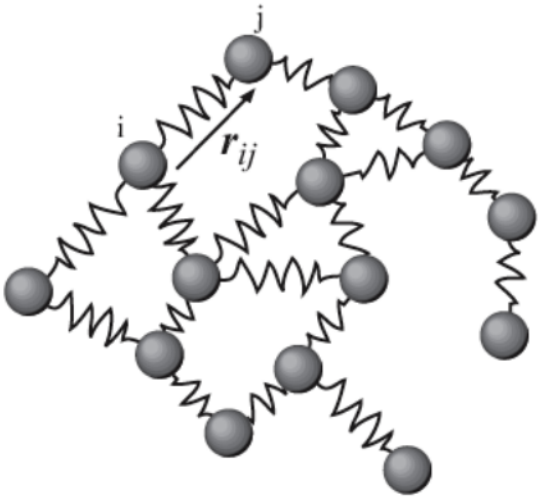

Figure 4: Spring System

4 Reconciliation: SpringRank

SpringRank is a ranking method based on a physical system of springs. (See figure 4) In the system, each small light ball represents a vertex, and each spring connecting two balls represents the edge between the vertices. The ranking score of a vertex depends on its energy. The energy between the two vertices can be represented by this function: [3]

$$H_{ij} = \frac{1}{2}(s_i - s_j - 1)^2 \tag{6}$$

, where *si–s j* represents the distance between vertex i and j. It shows that we can change the energy of the system by changing the distance between vertices. Therefore, by summing up the partial energies, we can get the total energy of the system: [3]

$$H_{ij} = \sum_{i,j=1}^{N} A_{ij} H_{ij} \quad = \frac{1}{2} \sum_{i,j=1}^{N} A_{ij}(s_i - s_j - 1)^2 \tag{7}$$

When we minimize the total energy of the system, we can get the most stable status of the system, from which we get the optimal rankings of the vertices. With Boltzmann distribution, we can get that the probability of i beating j has the following distribution function: [4]

$$\begin{aligned} P(i \to j) &= \frac{e^{-\beta H_{ij}}}{e^{-\beta H_{ij}} + e^{-\beta H_{ji}}} \\ &= \frac{1}{1 + e^{-2\beta(s_i - s_j)}} \end{aligned} \tag{8}$$

Which turns out to be another logistic distribution! Compared to the BTL model (equation 5), we can see that they have very similar structures, and the probabilities all depends on difference. Since the ranking score depends on energy of the system, it has some actual meanings, and thus is likely to represent the utility in DCM.

5 Conclusion

In this program, I learned the basic concepts and properties of Discrete Choice Model and Networks, looked into the ranking theories, and finally find the reconciliation between networks and choice modeling – the SpringRank. My mentor and I also tried hard to explain all the ideas in real-life examples to dig their realistic meanings. However, we still need more research to explore the relationship between SpringRank and DCM. Moreover, as the old saying goes, "all roads lead to Rome". SpringRank might not be the only way of connecting networks and choice modeling. Is there any other ranking method that works even better than SpringRank? Is there still any connection between networks and choice modeling if the choice model is continuous? There are still way more questions to answer and a long way to go in this young field.

References

[1] Kenneth Train. "Discrete Choice Methods with Simulation." March 8, 2002

[2] Clauset and Larremore. "Network Analysis and Modeling, CSCI 5352 Lecture Notes."

[3] Caterina De Bacco, Daniel B. Larremore, and Christopher Moore. "A Physical Model for Efficient Ranking in Networks."

[4] Aparajithan Venkateswaran. "Understanding SpringRank through Random Utility Models, Identifiability, and Online Updates." 2020