Binary Search Approach for Largest Cascade Capacity of Complex Networks

Natarajan Meghanathan
Jackson State University, USA

INTRODUCTION

Information Cascade in network theory is a behavioral phenomenon by which every node arrives at a decision (adopt or reject a particular thing) under the influence of the decision taken by its neighbor nodes (Easley & Kleinberg, 2010). The decision could be with regards to anything like adopting a new technology, supporting a particular political party or leader, choosing a lawn maintenance company, eating in a restaurant, etc. Given an initial set of nodes (called the initial adopters), the phenomena of information cascade (Easley & Kleinberg, 2010) goes through a series of iterations in each of which at least one node (that has not yet taken a decision) takes the decision under the influence of the decision taken by its neighbor nodes. The iterations stop when all the nodes have taken a decision or when no new node (i.e., no node other than those who had decided in the previous iterations) takes a decision in an iteration.

We refer to the information cascade as a *complete* information cascade (Easley & Kleinberg, 2010) if all the nodes arrive at a *unanimous* decision (for example: all nodes in a social network decide to support a particular political party in an upcoming election). For complete information cascade to happen, nodes (even if they have their own opinion) are expected to get influenced by the decision of their neighbor nodes so that the decision is eventually unanimous when the iterations stop. We assume a node will be in a position to take the *unanimous* decision when at least a *threshold fraction* of its neighbor nodes have taken/adopted the same decision (Easley & Kleinberg, 2010). For a given set of initial adopters, the maximum value for such a threshold fraction of adopted neighbor nodes in every neighborhood that can eventually enforce a unanimous decision for the entire network is called the *cascade capacity* of the network (Easley & Kleinberg, 2010).

The current approach (Easley & Kleinberg, 2010) used to determine the cascade capacity of a network is to first determine the clusters of the network and then determine their densities. The density of a cluster is the minimum of the intra cluster density (fraction of the incident edges to nodes within the same cluster) of the bridge nodes of the cluster (nodes that have one or more edges to nodes in other clusters). Information cascade can penetrate to a cluster and be complete only if the threshold fraction of adopted neighbor nodes (needed for adopting a unanimous decision) is less than or equal to 1 - the cluster density (Easley & Kleinberg, 2010). The cascade capacity of a network is the minimum of such threshold fractions of adopted neighbor nodes needed for a unanimous decision (Easley & Kleinberg, 2010). A primary weakness with the above approach is that it does not consider the nodes chosen as initial adopters to kick start information cascade while determining the cascade capacity of the network (also reported by Chesney, 2017). We claim that the above approach only gives a lower bound for the cascade capacity of the network and the cascade capacity of the network could be indeed larger if the initial adopters are also considered. Moreover, the above approach is time consuming as it first requires

DOI: 10.4018/978-1-7998-9220-5.ch150

to identify the clusters of a network and then identify the bridge nodes per cluster as well as determine their intra cluster density.

Our hypothesis for this research is that the cascade capacity of a network depends on the number and topological positions of the nodes chosen as initial adopters. To validate our hypothesis, we propose a binary search algorithm to determine the largest possible cascade capacity of a network for a given set of initial adopters. The binary search algorithm is briefly explained here: The search space (range of possible values for the threshold fraction of adopted neighbors for a unanimous decision) of the algorithm ranges from 0 (initial left index) to 1 (initial right index). We maintain an invariant that the information cascade will be complete if the left index is used as the threshold fraction of adopted neighbors and that the information cascade will not be complete if the right index is used as the threshold fraction of adopted neighbors. In each iteration, we find the middle index (average of the left and right index) and check if the information cascade can be complete when the middle index is used as the threshold fraction of adopted neighbors: if the information cascade is complete, we move the left index to the right and set the current value of the middle index to be the latest value of the left index; otherwise, we move the right index to the left and set the current value of the middle index to be the latest value of the right index. We proceed as long as the difference between the latest values of the left index and right index stays greater than or equal to a termination threshold (ε). Once the difference between the left index and right index becomes less than the termination threshold, we stop the algorithm and return the latest value of the left index as the largest possible cascade capacity of the network for the given set of initial adopters. The number of iterations needed by the binary search algorithm is $\log_2(1/\varepsilon)$.

Centrality metrics quantify the topological importance of the nodes in a network and are typically neighborhood-based or shortest path-based (Newman, 2010). We consider the neighborhood-based degree (DEG) and eigenvector (EVC) centrality metrics and the shortest path-based betweenness (BWC) and closeness (CLC) centrality metrics for our analysis. The degree centrality (Newman, 2010) of a node is the number of neighbors of the node. The eigenvector centrality (Bonacich, 1987) of a node is a measure of the degree of the node as well as the degrees of the neighbors of the node. The betweenness centrality (Freeman, 1977) of a node is a measure of the fractions of the shortest paths between any two nodes (in the network) that go through the node. The closeness centrality (Freeman, 1979) of a node is a measure of the distance (number of edges on the shortest path) of the node to the rest of the nodes in the network. After running the proposed binary search algorithm on a suite of 60 real-world networks, we observe the DEG centrality metric, followed by the BWC metric, to be relatively more effective for accomplishing larger cascade capacities for the networks when operated with the different percentages of initial adopters.

The rest of the chapter is organized as follows: The Background section reviews related work in the literature and highlights the unique contributions of this work. The section titled "Hypothesis and Motivating Example" presents our hypothesis and a motivating example to illustrate the impact of initial adopters on the cascade capacity of a network. The section titled "Iterative Algorithm and Information Cascade" presents an iterative algorithm used in this chapter to conduct information cascade in a network for a given set of initial adopters and threshold fraction of adopted neighbors. The section titled "Binary Search Algorithm" presents the proposed binary search algorithm to determine the cascade capacity of a network for a given set of initial adopters. The section titled "Analysis of Real-World Networks" presents the results (largest possible cascade capacities for a suite of 60 real-world networks for 2%, 5% and 10% of the nodes as initial adopters with respect to each of the four centrality metrics) and analyzes the increase in the cascade capacities of the networks with increase in the % of the initial adopters vis-a-vis those determined using the intra cluster density approach. The section titled "Future Research Directions"

Related Content