Maximizing Diversity of Opinion in Social Networks

Erika Mackin and Stacy Patterson

Abstract—We study the problem of maximizing opinion diversity in a social network that includes opinion leaders with binary opposing opinions. The members of the network who are not leaders form their opinions using the French-DeGroot model of opinion dynamics. To quantify the diversity of such a system, we adapt two diversity measures from ecology to our setting, the Simpson Diversity Index and the Shannon Index. Using these two measures, we formalize the problem of how to place a single leader with opinion 1, given a network with a leader with opinion 0, so as to maximize the opinion diversity. We give analytical solutions to these problems for paths, cycles, and trees, and we highlight our results through a numerical example.

I. INTRODUCTION

As social networks and social media become an increasingly important part of our lives, the study of how opinions form and spread in these networks has become a rich area of study. Problems of interest include how quickly members of a social network converge to agreement [1], how to identify the most influential users [2], the effect of friendly vs. antagonistic interactions between users [3], which users to advertise to in order to increase the popularity of a product [4], and so on.

One potential downside of online social media spaces is that they encourage rapid consensus and polarization of opinions [5], [6]. Although agreement in a population has its benefits, diversity of opinion is also important. A community with diverse opinions is better able to innovate due to a wider variety of potential perspectives [7]. Further, it has been argued that there are four traits necessary for a crowd to be wise, one of which is diversity of opinion [8].

We study a social network where members, or nodes, exchange opinions using the French-DeGroot model, wherein nodes update their opinion based on both their current opinion and the opinions of their peers [9]. We assume the system contains some nodes who contribute their opinions to their neighbors but never change their own opinions. In real-world social networks, such a person could be a paid promoter of some product or political stance (i.e., social media “influencers”), or simply very opinionated. These nodes are the opinion leaders in the network, which we refer to simply as leaders. We consider a setting in which leaders each have an opinion of 0 or 1. For example, in an election, perhaps an opinion of 0.5 indicating a truly undecided voter. In this setting, the follower nodes’ opinions converge to values in the interval [0, 1]. Our aim is to quantify the diversity of these opinions; intuitively, a diverse network has opinions that cover the full spectrum of the opinion interval [0, 1].

To formalize this notion of diversity, we propose two diversity measures: the Simpson Opinion Diversity Index and the Shannon Opinion Diversity Index. The first is based on the Simpson Diversity Index, a concept originally from ecology that measures the diversity of a community composed of many different species [10]. The Simpson Diversity Index was designed to be maximized when all species in the area are represented equally. Our second measure is derived from the Shannon Index, which was originally developed in a communications context to express the probability of receiving a given text string over a communication channel. In this context, it is now generally referred to as information entropy. The Shannon Index is also used in ecology, where the probabilities represent the likelihood of a randomly selected individual belonging to a given species [11]. The Shannon Index is maximized when all outcomes are equally probable.

Our proposed performance measures are parameterized by a number of opinion bins. This can range from two bins, where a maximally diverse network is one in which half the population has an opinion in the interval [0, 0.5] and the remaining population has an opinion in the interval [0.5, 1], to \( n_f \) bins, in a network with \( n_f \) follower nodes, in which case a maximally diverse network is one in which the opinions are uniformly distributed over [0, 1].

We pose the problem of optimizing the diversity of the network by selecting which nodes should act as leaders. Specifically, we consider a network with a single leader node with opinion 0. As is, such a network will converge to a state where all opinions are 0. We seek to identify a node that, if it becomes a 1-valued leader, will maximize the diversity of the resulting opinions in the network. For instance, if we want to prevent political discussion in an online forum from being dominated by a supporter of Party A, we can invite a strong supporter of Party B into the community as well to encourage a wider variety of opinions. We present analytical solutions to this problem for both two bins and \( n_f \) bins in path, cycle, and tree graphs for both performance measures. Further, we present numerical results highlighting the difference between these performance measures.
Related Work: Previous work on controlling the opinions of the network has generally focused on maximizing either the sum or the average of the node states, rather than promoting diverse opinions. Targeted placement of leaders to maximize the followers’ opinions has been studied in systems with stubborn agents [12], agents with both fixed internal and modifiable external opinions [13], agents whose stubbornness increases over time [14], and opposing leaders with similar dynamics to our setting [15]. In [16], they consider a network that consists of binary opposing leaders and followers who update their state via binary voting and study the problem of maximizing the sum of the node states. And in [17], the problem of how to place a single leader in the problem of maximizing the sum of the node states. The dynamics for each follower can be expressed more compactly as

\[
\dot{x}_f(t) = -L_{ff} x_f - L_{fl} x_l.
\]

It has been shown that, under these dynamics, the followers’ states converge to a convex combination of the leader states, which is given by the following expression [18]

\[
\dot{x}_f = -L_{ff}^{-1} L_{fl} x_l.
\]

We call \( x_f \) the opinion vector of the network \( G \). We note that since all leader opinions are either 0 or 1, for each follower \( v \), \( x_{f_v} \in \{0, 1\} \).

A. Diversity Performance Measures

We quantify the diversity of the opinion vector \( x_f \) using two different diversity measures. Our first measure is based on the Simpson Diversity Index. This index was originally introduced as a measure of biological diversity, where a region with an even distribution of species is considered to be more diverse than an area where the population is dominated by only a few types of organisms.

Definition 1: Consider a region with \( R \) species, where each species \( i \) has \( n_i \) members present in the area. The Simpson Diversity Index is [10]:

\[
SDI = 1 - \sum_{i=1}^{R} \frac{n_i(n_i - 1)}{n(n - 1)}
\]

Under this definition, \( SDI = 1 \) represents infinite diversity, and \( SDI = 0 \) indicates complete domination by a single species or category.

We adopt this measure to the opinion vector by first discretizing the interval \([0, 1]\) into \( R \) bins:

\[
b_{i} = \frac{i - 1}{R}, \quad i = 2 \ldots (R - 1) \tag{2}
\]

\[
b_{R} = \frac{R - 1}{R}, \tag{3}
\]

We then count the number of opinions in \( \hat{x}_f \) that fall into each bin. The opinion diversity is measured as follows.
Definition 2: Let $c_i$ be the number of components of $\hat{x}_f$ that lie in bin $b_i$. The Simpson Opinion Diversity Index of a network $G$ with 0-leader set $S_0$ and 1-leader set $S_1$ is:

$$O_{\text{Sim}}(S_0, S_1, R) = 1 - \sum_{i=1}^{R} \frac{c_i}{n_f} \left( 1 - \frac{c_i}{n_f} \right)$$

(4)

Our second diversity measure is based on the Shannon Index, a measure that was developed to quantify the entropy in a text string. This index has also been applied to measuring diversity in ecosystems.

Definition 3: Consider a region with $R$ species, where each species $i$ has $n_i$ members present in the area. The Shannon Index is [11]:

$$SI = -\sum_{i=1}^{R} p_i \ln(p_i)$$

where $p_i$ is the number of individuals that belong to category $i$ divided by the number of categories.

Note that this index is maximized when the individuals are evenly distributed among all $R$ categories, and is always non-negative since $p_i \in [0, 1]$.

To adapt this measure to the opinion vector, we follow a similar procedure as for the Simpson Opinion Diversity Index. We divide the interval into $R$ bins as shown in (1) - (3) and use these bins in our definition.

Definition 4: Let $p_i$ be the proportion of components of $\hat{x}_f$ that lie in bin $b_i$, i.e., $p_i = c_i/R$, with $p_i \ln(p_i) = 0$ when $p_i = 0$. The Shannon Opinion Diversity Index of a network $G$ with 0-leader set $S_0$ and 1-leader set $S_1$ is:

$$O_{\text{Shan}}(S_0, S_1, R) = -\sum_{i=1}^{R} p_i \ln(p_i)$$

When there are $n_f$ buckets and there exist optimal leader sets $S_0^*, S_1^*$ so that all $n_f$ follower opinions are uniformly distributed, then $c_i = 1$ for $i = 1, \ldots, n_f$. It follows that

$$O_{\text{Sim}}(S_0^*, S_1^*, n_f) = 1$$

(5)

and

$$O_{\text{Shan}}(S_0^*, S_1^*, n_f) = -\ln \left( \frac{1}{n_f} \right)$$

(6)

are the maximum values for both measures. Note, as $n_f$ increases, the maximum value of $O_{\text{Shan}}$ is unbounded.

When $R = 2$ and the leader sets $S_0^*, S_1^*$ are selected so that all $n_f$ followers are uniformly distributed, then $|c_1 - c_2| \leq 1$. Without loss of generality, let $c_1 = \left\lfloor \frac{n_f}{2} \right\rfloor$ and $c_2 = \left\lceil \frac{n_f}{2} \right\rceil$. Then

$$O_{\text{Sim}}(S_0^*, S_1^*, n_f) = 1 - \frac{n_f}{2(n_f - 1)} - \frac{n_f(n_f - 1)}{n_f(n_f - 1)}$$

(7)

and

$$O_{\text{Shan}}(S_0^*, S_1^*, n_f) = -\frac{n_f}{2} \ln \left( \frac{n_f}{2} \right) - \frac{n_f}{2} \ln \left( \frac{n_f}{2} \right)$$

(8)

are the maximum values for both measures.

In some networks, even when $S_0^*, S_1^*$ are chosen as the leader sets, uniform distribution of follower opinions is unachievable. In such a case, the above expressions serve as an upper bound on the resulting diversity indices.

B. Problem Formulation

Let the pre-existing 0-leader be denoted by $l_0$. We want to determine where to place the 1-leader, $l_1$, such that the diversity of opinions present in $x_f$ is maximized. We pose two optimization problems, one for each diversity measure.

The Leader selection problem for Simpson Opinion Diversity is:

$$\max_{l_1} \ N_{\text{Sim}}(l_0, l_1, R)$$

(9)

The Leader selection problem for Shannon Opinion Diversity is:

$$\max_{l_1} \ N_{\text{Shan}}(l_0, l_1, R)$$

(10)

III. ANALYSIS

We now present analytical solutions to (LS1) and (LS2) for $K = 1$ in several classes of graphs for $R = n_f$ and $R = 2$. Since $K = 1$, we call the single 0-leader $l_0$ and the 1-leader $l_1$.

A. Path Graphs

We first consider a path graph of $n$ nodes, numbered $1, 2, \ldots, n$. The following theorem addresses the case where $R = n_f$ for both diversity indices and proves that the optimal placement of $l_1$ is at the farthest node from $l_0$.

Theorem 1: Consider a path of length $n$, with a single 0-leader node $k$. The optimal solution to both (LS1) and (LS2), for $R = n_f$, is to select node $j$ as the single 1-leader node, where $j = n$ if $k < n/2$ and $j = 1$ otherwise.

To prove this theorem, first we note the useful fact that, for a path graph with $l_0 = k$ and $l_1 = j$, where $k < j$ [15],

$$\hat{x}_{f_o} = \frac{v - k}{j - k} \text{ for } v = k + 1, \ldots, j - 1.$$
Proof: Without loss of generality, assume \( l_0 = k \) with \( k < \frac{n}{2} \). Further, assume \( l_1 = j \), with \( j > k \). We can then note that there are \( k-1 \) nodes with opinion 0, and \( n-j \) nodes with opinion 1. Thus, \( c_1 \geq k-1 \) and \( c_{nj} \geq n-j \). Note that all bins have width \( \frac{1}{nf} \). When the number of followers between \( l_0 \) and \( l_1 \) is \( z < nf \), the distance between the opinions of each consecutive pair of followers is \( \frac{1}{2} < \frac{1}{nf} \) by (9). Therefore, each bin \( b_i \), \( i = 2, \ldots, n_f-1 \), must have \( c_i \leq 1 \). Consider two cases:

Case 1: Let \( j-k = n_f \), so that \( k = n-j \). Then by Lemma 1, \( c_i = 1 \) for all \( i = 1, \ldots, n_f \) and \( O_{Sim}(1, n, n_f) = 1 \). When \( l_1 \) is node \( n-1 \), then by Lemma 2, \( c_i = 0 \), \( c_{nj} = 2 \), and \( c_j = 1 \) for all \( j = 2, \ldots, n_f-1 \). Then

\[
O_{Sim}(1, n-1, n_f) = 1 - \frac{2}{n_f(n_f-1)}.
\]

Clearly, \( O_{Sim}(1, n, n_f) - O_{Sim}(1, n-1, n_f) > 0 \), and the follower states are more diverse when \( l_1 \) is \( n \) than when \( l_1 \) is \( n-1 \).

Case 2: Let \( j < n \) so that \( j-k < n_f \). Then \( c_i = k-1 \) and \( c_{nj} = g \) where \( g \geq n-j \). When \( l_1 \) is node \( j-1 \), the difference between the opinions of follower nodes \( i \) and \( i+1 \), where \( i = 2, \ldots, j-2 \), increases from \( \frac{1}{k-1} \) to \( \frac{1}{j-k} \) by (9). Therefore, bins \( i = 2, \ldots, n_f-1 \) have count \( c_i \leq 1 \). The count of bin \( b_{nj} \) increases to \( g+1 \), since node \( j \) now has opinion 1 as well, while the count of bin \( b_1 \) is unchanged.

Then \( O_{Sim} \) is computed as follows:

\[
O_{Sim}(1, j, n_f) = 1 - \frac{(k-1)(k-2)}{n_f(n_f-1)} - \frac{g(g-1)}{n_f(n_f)}.
\]

and

\[
O_{Sim}(1, j-1, n_f) = 1 - \frac{(k-1)(k-2)}{n_f(n_f-1)} - \frac{(g+1)g}{n_f(n_f)}.
\]

The difference is then

\[
O_{Sim}(1, j, n_f) - O_{Sim}(1, j-1, n_f) = \frac{2g}{n_f(n_f-1)} > 0;
\]

therefore, moving \( l_1 \) from \( j \) to \( j-1 \) decreases \( O_{Sim} \).

We use these same two cases and their associated bin counts for \( O_{Shan} \). In Case 1:

\[
O_{Shan}(1, n, n_f) = -n_f \frac{1}{nf} \ln \left( \frac{1}{n_f} \right)
\]

and

\[
O_{Shan}(1, n-1, n_f) = -(n_f-2) \frac{1}{nf} \ln \left( \frac{1}{n_f} \right) - \frac{2}{nf} \ln \left( \frac{2}{n_f} \right)
\]

and their difference is:

\[
O_{Shan}(1, n, n_f) - O_{Shan}(1, n-1, n_f) = \frac{2}{nf} \ln 2 > 0.
\]

In Case 2:

\[
O_{Shan}(k, j, n_f) = -\frac{k-1}{n_f} \ln \left( \frac{k-1}{n_f} \right)
\]

\[
- \frac{g}{n_f} \ln \left( \frac{g}{n_f} \right) - (n_f-g-k+1) \frac{1}{n_f} \ln \left( \frac{1}{n_f} \right)
\]

and

\[
O_{Shan}(k, j-1, n_f) = -\frac{k-1}{n_f} \ln \left( \frac{k-1}{n_f} \right)
\]

\[
- \frac{g+1}{n_f} \ln \left( \frac{g+1}{n_f} \right) - (n_f-g-k) \frac{1}{n_f} \ln \left( \frac{1}{n_f} \right)
\]

and their difference is:

\[
O_{Shan}(k, j, n_f) - O_{Shan}(k, j-1, n_f)
\]

\[
= \frac{g}{n_f} \ln \left( \frac{g}{n_f} \right) + \frac{1}{n_f} \ln (g+1) > 0.
\]

We can see that in both cases, moving \( l_1 \) one node closer to \( l_0 \) decreases \( O_{Shan} \). Thus, for both problems and both cases, moving \( l_1 \) closer to \( l_0 \) always decreases the diversity, and therefore, placing \( l_1 \) at node \( n \) is optimal.

A similar argument can be used to show that it is suboptimal to select \( j < k \) for \( l_1 \) for both \( O_{Sim} \) and \( O_{Shan} \).

By Theorem 1, when \( l_0 = k < n/2 \), the optimal 1-leader is \( l_1 = n \) for both diversity indices. The resulting diversity measures are:

\[
O_{Sim}(k, n, n_f) = 1 - \frac{(k-1)(k-2)}{n_f(n_f-1)}
\]

\[
O_{Shan}(k, n, n_f) = -\frac{k-1}{n_f} \ln \left( \frac{k-1}{n_f} \right) - \frac{n_f-(k-1)}{n_f} \ln \left( \frac{1}{n_f} \right).
\]

Thus, even when \( l_1 \) is chosen optimally, the diversity may be far from the maximal diversity value of 1 or \(-\ln \left( \frac{1}{n_f} \right) \), for (LS1) and (LS2), respectively. And, these indices are farthest from their maximal value when \( l_0 = (n/2) - 1 \).

We now consider the maximization of Problems (LS1) and (LS2) when \( R = 2 \) and prove that, unlike when \( R = N_f \), diversity is maximized when \( l_1 \) is chosen so that \( l_0 \) and \( l_1 \) are the same distance from the endpoints.

Theorem 2: Consider a path of \( n \) nodes, with a single 0-leader \( l_0 = k \). An optimal solution to both (LS1) and (LS2), for \( R = 2 \), is to select node \( j \) as the single 1-leader, where \( j = n-k+1 \) when \( k < \frac{n}{2} \) and \( j = n-k \) otherwise.

Proof: Note that for \( R = 2 \), the indices are computed simply as:

\[
O_{Sim}(S_0, S_1, B_2) = 1 - \frac{c_1(c_1-1)}{n_f(n_f-1)} - \frac{c_2(c_2-1)}{n_f(n_f-1)}
\]

and

\[
O_{Shan}(S_0, S_1, B_2) = -\frac{c_1}{n_f} \ln \left( \frac{c_1}{n_f} \right) - \frac{c_2}{n_f} \ln \left( \frac{c_2}{n_f} \right),
\]

and that, by definition, both \( O_{Sim} \) and \( O_{Shan} \) are maximized when the opinions of the \( n_f \) followers are evenly distributed between the two bins.

Without loss of generality, assume \( k < \frac{n}{2} \). There are two cases:
Case 1: Let $n_f$ be even. Observe that when $l_1 = j = n - k - 1$, there are $n_f - 2(k - 1)$ follower nodes that fall between $l_0$ and $l_1$. Then $c_1 = k - 1 + \frac{n_f - 2(k - 1)}{2} = \frac{n_f - 2}{2} = c_2$.

Case 2: Let $n_f$ be odd. Then it is impossible for $c_1$ and $c_2$ to be equal. Let $l_1$ be node $j = n - k + 1$. Then once again there are $n_f - 2(k - 1)$ follower nodes between $l_0$ and $l_1$, and so $c_1 = k - 1 + \frac{n_f - 2(k - 1) - 1}{2} = \frac{n_f - 1}{2}$ and $c_2 = k - 1 + \frac{n_f - 2(k - 1) + 1}{2} = \frac{n_f + 1}{2}$ and thus $c_1 = c_2 - 1$.

In both cases, both diversity indices are maximized when $l_1$ is node $j = n - k + 1$.

A similar argument can be used to show that both diversity indices are maximized when $l_1$ is node $n - k$ when $k > \frac{n}{2}$.

Unlike the case when $R = n_f$, when $R = 2$ and $n_f$ is even, regardless of which node is $l_0$, it is always possible to find an $l_1$ such the resulting opinion diversity is maximal, as given in (7) and (8).

B. Cycle graphs

We now consider (LS1) and (LS2), when $R = n_f$, over a cycle of $n$ nodes, numbered $1, 2, \ldots, n$, in a clockwise manner. We assume, without loss of generality, that $l_0$ is node 1. In such a setting, we prove that diversity is maximized when $l_1$ is placed directly beside $l_0$.

Theorem 3: Consider a cycle of $n$ nodes with 0-leader node $l_0 = 1$, and let $R = n_f$. The optimal solutions to Problems (LS1) and (LS2) are $l_1 = 2$ and $l_1 = n - 1$, respectively.

Proof: Note that when $l_1$ is either node 2 or $n - 1$, the follower states $x_f$ are the same as when the graph is a path of length $n$ with $l_0$ and $l_1$ located at the endpoints. Thus, by the same argument as in the proof of Theorem 1, $c_i = 1$ for all $i = 1, \ldots, n_f$. When $l_1$ is at node $u, u \neq 2, n - 1$, the cycle is broken into two paths, with one path having $p$ nodes between $l_0$ and $l_1$ and the other having $n_f - p$ nodes (without loss of generality, assume $p \leq n_f - p$). Since both resulting paths have a length less than $n$, by (9) we note that $c_1 = 0$ for both paths and $c_i \geq 2$ for some $i \neq 1$. Thus,

$$O_{Sim}(1, 2, n_f) = O_{Sim}(1, n - 1, n_f) > O_{Sim}(1, u, n_f)$$

and

$$O_{Shan}(1, 2, n_f) = O_{Shan}(1, n - 1, n_f) > O_{Shan}(1, u, n_f).$$

Therefore, the optimal $l_1$’s are 2 and $n - 1$.

When $R = n_f$, $l_0 = 1$ and an optimal 1-leader $l_1 = 2$ or $l_1 = n - 1$ is selected, then

$$O_{Sim}(1, l_1, n_f) = 1$$

and

$$O_{Shan}(1, l_1, n_f) = -\ln \left( \frac{1}{n_f} \right).$$

Clearly, $O_{Sim}(1, l_1, n_f) = O_{Sim}(S_0, S_1, n_f)$ and $O_{Shan}(1, l_1, n_f) = O_{Shan}(S_0, S_1, n_f)$. Therefore, when $R = n_f$ and the graph is a cycle, the maximal diversity can be achieved, regardless of the location of $l_0$.

Next, we present results for the case where $R = 2$.

Theorem 4: Consider a cycle of $n$ nodes with 0-leader node $l_0 = 1$. When $R = 2$, and $n_f$ is odd, $l_1 = j$ is an optimal solution for all $j = 2, \ldots, n$ for both Problems (LS1) and (LS2). When $n_f$ is even, $l_1 = j$ is an optimal solution for all $j = 2, 4, \ldots, n$ for both (LS1) and (LS2).

Proof: Assume that $l_1 = j$, such that the cycle is broken into two paths. Let one path have $p \geq 0$ nodes between $l_0$ and $l_1$ and the other have $n_f - p$. Without loss of generality, assume $p \leq n_f - p$.

We once again note that, in this setting,

$$O_{Sim}(S_0, S_1, B_2) = 1 - \frac{c_1(c_1 - 1)}{n_f(n_f - 1)} - \frac{c_2(c_2 - 2)}{n_f(n_f - 1)}$$

and

$$O_{Shan}(S_0, S_1, B_2) = -\frac{c_1}{n_f} \ln \left( \frac{c_1}{n_f} \right) - \frac{c_2}{n_f} \ln \left( \frac{c_2}{n_f} \right).$$

By definition, both $O_{Sim}$ and $O_{Shan}$ are maximized when $n_f$ is evenly distributed between the two bins. We consider four cases and show that in each the location of $l_1$ has no effect on the bin counts $c_1$ and $c_2$.

Case 1: Let $n_f$ be odd and $p$ be odd. Then $n_f - p$ must be even. Then $c_1 = \frac{p+1}{2} + \frac{n_f-p}{2} = \frac{n_f-1}{2}$ and $c_2 = \frac{p+1}{2} + \frac{n_f-p}{2} = \frac{n_f+1}{2}$.

Case 2: Let $n_f$ be odd and $p$ be even so that $n_f - p$ must be odd. Then $c_1 = \frac{p}{2} + \frac{n_f-p-1}{2} = \frac{n_f-2}{2}$ and $c_2 = \frac{p}{2} + \frac{n_f-p+1}{2} = \frac{n_f+2}{2}$.

Case 3: Let $n_f$ be even and $p$ be odd, so $n_f - p$ is also odd. Then $c_1 = \frac{p-1}{2} + \frac{n_f-p-1}{2} = \frac{n_f-2}{2}$ and $c_2 = \frac{p+1}{2} + \frac{n_f-p+1}{2} = \frac{n_f+2}{2}$.

Case 4: Finally, let $n_f$ be even, so that both $p$ and $n_f - p$ are even. Then $c_1 = \frac{p}{2} + \frac{n_f-p}{2} = \frac{n_f-1}{2}$ and $c_2 = \frac{p}{2} + \frac{n_f-p}{2} = \frac{n_f+1}{2}$.

In all four cases, $c_1$ and $c_2$ are independent of $p$ and $n_f - p$. When $n_f$ is odd, all possible locations of $l_1$ are equivalent and have $|c_1 - c_2| = 1$, and, thus, all $l_1 = j, j = 2, \ldots, n$ optimize both performance measures. When $n_f$ is even, $|c_1 - c_2| \leq 1$ only when $p$ and $n_f - p$ are both even. Therefore $l_1 = j$ is equivalent for all $j = 2, 4, \ldots, n$, and all such $l_1 = j$ optimize both performance measures.

When $R = 2$, the graph is a cycle, $l_0 = 1$, and $n_f$ is odd, then all possible placements of $l_1$ are optimal. In the case where $n_f$ is even, $l_1 = j$ is optimal only when $j$ is even.

C. Tree graphs

Finally, we consider (LS1) and (LS2) over tree graphs. We first study the problem when $R = n_f$ in a special class of tree graph.

Theorem 5: Consider a tree graph of size $n$ where exactly one node $i$ has degree $deg(i) = 3$ and all other nodes $j$ have $deg(j) \leq 2$. Let the 0-leader $l_0$ be at a leaf node. Then Problems (LS1) and (LS2) are both maximized when $l_1$ is at the end of the longest path from $l_0$.

The proof of this theorem is quite lengthy and is deferred to a technical report [19].
Fig. 1: Example of a tree network $G$, where, when $R = 2$, $l_0$ and $l_1$ are optimal leaders, satisfying Theorem 6.

We also consider the problem of identifying an optimal $l_1$ in a general tree graph when $R = 2$.

Note that the set of followers in any graph with one 0-leader and one 1-leader can be partitioned into three sets, $P_1$, $P_2$, and $P_3$, based on the locations of the leaders. Let $\text{path}(a, b)$ be defined on a tree graph as the set of all nodes that lie on the path between nodes $a$ and $b$, inclusive of $a$ and $b$. Then, the three sets are defined as follows:

- $P_1 = \{i \in V : l_0 \in \text{path}(i, l_1)\}$
- $P_2 = \{i \in V : l_1 \notin \text{path}(i, l_0) \text{ and } l_0 \notin \text{path}(i, l_1)\}$
- $P_3 = \{i \in V : l_1 \in \text{path}(i, l_0)\}$

where $|P_1 \cup P_2 \cup P_3| = n_f$ and $P_1 \cap P_2 = P_1 \cap P_3 = P_2 \cap P_3 = 0$. Note also that, in a tree, the graph distance between two nodes, $d(u, v)$, is the length of the path between nodes $u$ and $v$. In graph $G$, shown in Figure 1, there are 10 follower nodes. Given $l_0$, $l_1$ as shown, $P_1 = \{1, 2, 3\}$, $P_2 = \{4, 5, 6, 7\}$, and $P_3 = \{8, 9, 10\}$.

We now consider the case when $R = 2$ and $l_0$ and $l_1$ are placed such that $|P_1| = |P_3|$, that is, the number of followers with opinion 0 and opinion 1 are the same. If the network is such that the nodes that lie between the leaders have evenly distributed opinions, then the existing leader placement is optimal, regardless of the size of $|P_1|$ and $P_3$. An example network showing such a leader placement is given in Figure 1.

Theorem 6: Consider a tree graph $G$ with $n$ nodes, where $l_0 = i$ and $|P_1| = k - 1$, and let $R = 2$. Without loss of generality, let $k < \frac{n}{2}$, so that $|P_2 \cup P_3| = n_f - (k - 1)$. If there exists a node $j$ such that, when $l_1 = j$, $|P_3| = k - 1$, $|P_2| = n_f - 2(k - 1)$, and the opinions of nodes $i \in P_2$ are evenly distributed between bins $b_1$ and $b_2$, then $j$ is an $l_1$ that maximizes both (LS1) and (LS2).

Proof: We once again note that, when $R = 2$, that

$$O_{\text{Sim}}(I_0, j, B_2) = 1 - \frac{c_1(c_1 - 1)}{n_f(n_f - 1)} - \frac{c_2(c_2 - 1)}{n_f(n_f - 1)}$$

and

$$O_{\text{Shan}}(I_0, j, B_2) = -\frac{c_1 n_f}{n_f} \ln \left(\frac{c_1}{n_f}\right) - \frac{c_2 n_f}{n_f} \ln \left(\frac{c_2}{n_f}\right).$$

By definition, both $O_{\text{Sim}}$ and $O_{\text{Shan}}$ are maximized when $n_f$ is evenly distributed between the two bins so that $|c_1 - c_2| \leq 1$.

There are two cases to consider.

Case 1: When $n_f$ is even, both (LS1) and (LS2) are maximized when $c_1 = c_2 = \frac{n_f}{2}$. Let $l_1 = j$ such that $|P_3| = k - 1$ and $|P_2| = n_f - 2(k - 1)$. To ensure that the opinions of nodes $i \in P_2$ are divided evenly among $b_1$ and $b_2$, there must exist an edge $(a, b) \in E$, $a, b \in P_2$ with the following properties: $b \in \text{path}(a, l_0)$, $a \in \text{path}(b, l_0)$, $d(l_0, a) = d(b, l_1)$, and

$$|\{i : b \in \text{path}(i, l_1)\}| = |\{j : a \in \text{path}(j, l_0)\}| = \frac{n_f - 2(k - 1)}{2}.$$

Then $c_1 = c_2 = \frac{n_f}{2}$, and $j$ is an optimal 1-leader.

Case 2: When $n_f$ is odd, both (LS1) and (LS2) are maximized when either

$$c_1 = \frac{n_f - 1}{2}, \quad c_2 = \frac{n_f + 1}{2}$$

or

$$c_1 = \frac{n_f + 1}{2}, \quad c_2 = \frac{n_f - 1}{2},$$

as in both cases $|c_1 - c_2| \leq 1$. Let $l_1 = j$ such that $|P_3| = k - 1$ and $|P_2| = n_f - 2(k - 1)$. To ensure that the opinions of nodes $i \in P_2$ are divided evenly among $b_1$ and $b_2$, there must exist a node $v \in P_2$ with the following properties:

$$d(l_0, v) = d(v, l_1),$$

$$|\{i : v \in \text{path}(i, l_1)\} \setminus \{v\}| = g_1,$$

$$|\{j : v \in \text{path}(j, l_0)\} \setminus \{v\}| = g_2.$$  

Note that, because $|P_2|$ is odd and $v$ is equidistant from $l_0$ and $l_1$, $v \in b_2$. Then the opinion diversity is maximized when either

$$g_1 = g_2 = \frac{n_f - 2(k - 1) - 1}{2}$$

and, thus, $c_1 = \frac{n_f - 1}{2}$ and $c_2 = \frac{n_f + 1}{2}$, or

$$g_1 = \frac{n_f - 2(k - 1) + 1}{2} \quad \text{and} \quad g_2 = \frac{n_f - 2(k - 1) - 3}{2},$$

in which case $c_1 = \frac{n_f + 1}{2}$ and $c_2 = \frac{n_f - 1}{2}$. For both pairs of $g_1$ and $g_2$, $|c_1 - c_2| = 1$ and $j$ is an optimal $l_1$.

Therefore, in both cases, when $l_1 = j$, $|c_1 - c_2| \leq 1$ and $l_1$ maximizes both (LS1) and (LS2) for $R = 2$. ■

Note that depending on the structure of $G$, such a partition of the follower nodes may not be possible. However, this does not imply that there are no other optimal choices of $l_1$.

IV. Numerical Examples

We highlight some of the analysis in Section III via numerical examples, using the graph $G$ shown in Fig. 2.

As shown in the previous section, for cycles and paths, the same $l_1$ node is optimal for both (LS1) and (LS2) for both $R = n_f$ and $R = 2$. We now show that, when $R = n_f$, this relationship between the optimal solutions does not always hold.

In the network $G$ in Fig. 2, when $l_0 = 1$, the optimal $l_1$ depends on the performance measure used. For (LS1), the optimal 1-leader can be either 10 or 11, but for (LS2) the optimal 1-leader is either 5 or 6.
TABLE I: Diversity indices for different $l_1$ values.

| $l_1$ | $O_{\text{Sim}}(1, l_1, n_f)$ | $O_{\text{Shan}}(1, l_1, n_f)$ |
|-------|-----------------------------|-----------------------------|
| 2     | 0                           | 0                           |
| 3     | 0.5                         | 0.637                       |
| 4     | 0.556                       | 0.849                       |
| 5     | 0.583                       | 1.003                       |
| 6     | 0.583                       | 1.003                       |
| 7     | 0.556                       | 0.687                       |
| 8     | 0.556                       | 0.687                       |
| 9     | 0.639                       | 0.937                       |
| 10    | 0.639                       | 0.937                       |
| 11    |                             |                             |

As shown in Table I, we can see that, for $y \in \{10,11\}$, $z \in \{5,6\}$,

$$O_{\text{Sim}}(1, y, n_f) = 0.639$$

$$> O_{\text{Sim}}(1, z, n_f) = 0.583$$

and

$$O_{\text{Shan}}(1, z, n_f) = 1.003$$

$$> O_{\text{Shan}}(1, y, n_f) = 0.937$$

We can use the same graph $G$ to observe that Theorem 5 does not generalize to trees with more than one node $i$ with degree $deg(i) \geq 3$. The longest path from $l_0$ in $G$ terminates at node 6, but, as shown above, 6 is not the optimal $l_1$ for Problem (LS1). Although determining the optimal $l_1$ is often simple, it is not a trivial problem. Although $R = n_f$, 11 and 6 are optimal $l_1$ nodes in $G$ for (LS1) and (LS2), respectively. The worst case $l_1$ node for both problems is 2, which results in $O_{\text{Sim}}(1,2, n_f) = 0$ and $O_{\text{Shan}}(1,2, n_f) = 0$.

Finally, we note that node 11 is an optimal $l_1$ node that satisfies the requirements listed in Theorem 6, such that $P_1 = P_3 = 0$, $P_2 = \{i : 2 \leq i \leq 10\}$, and $c_1 = \frac{n_f + 1}{2} = 5$ and $c_2 = \frac{n_f - 1}{2} = 4$. When $l_1 = 8$, 9, or 10, $c_1 = 5$ and $c_2 = 4$, but $|P_1| \neq |P_3|$. We can then see that there are optimal 1-leaders not described by Theorem 6.

V. CONCLUSION

We have proposed two diversity measures, adapted from ecology, for networks with French-DeGroot opinion dynamics, where the networks contain leaders with binary opposing opinions. Further, using these measures, we have formalized the problem of maximizing opinion diversity in a network that contains a single leader with opinion 0 by selecting which node should become the leader with opinion 1. We have presented analytical solutions to these problems for the case of a single 0-leader and a single 1-leader in paths, cycles, and tree graphs. In future work, we plan to study the problem of optimal leader placement in more general graphs and with multiple leaders of both opinion types.

REFERENCES

[1] J. Ghaderi and R. Srikant, “Opinion dynamics in social networks with stubborn agents: Equilibrium and convergence rate,” Automatica, vol. 50, no. 12, pp. 3209–3215, 2014.

[2] D. Kempe, J. Kleinberg, and É. Tardos, “Maximizing the spread of influence through a social network,” in Proc. 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2003, pp. 137–146.

[3] S. Zhou and W. X. Zheng, “Generalized dynamics in social networks with antagonistic interactions,” arXiv preprint arXiv:1801.08713, 2018.

[4] G. Toscani, A. Tosin, and M. Zanella, “Opinion modeling on social media and marketing aspects,” Physical Review E, vol. 98, no. 2, p. 022315, 2018.

[5] A. Das, S. Gollapudi, and K. Munagala, “Modeling opinion dynamics in social networks,” in Proc. 7th ACM International Conference on Web Search and Data Mining. ACM, 2014, pp. 403–412.

[6] P. Wallace, The psychology of the Internet. Cambridge University Press, 2015.

[7] S. E. Page, “Making the difference: Applying a logic of diversity,” Academy of Management Perspectives, vol. 21, no. 4, pp. 6–20, 2007.

[8] J. Surowiecki, The wisdom of crowds. Anchor, 2005.

[9] A. V. Proskurnikov and R. Tempo, “A tutorial on modeling and analysis of dynamic social networks. part ii,” Annual Reviews in Control, 2018.

[10] E. H. Simpson, “Measurement of diversity,” nature, 1949.

[11] C. E. Shannon, “A mathematical theory of communication,” ACM SIGMOBILE Mobile Computing and Communications Review, vol. 5, no. 1, pp. 3–55, 2001.

[12] D. S. Hunter and T. Zaman, “Opinion dynamics with stubborn agents,” arXiv preprint arXiv:1806.11255, 2018.

[13] A. Gionis, E. Terzi, and P. Tsaparas, “Opinion maximization in social networks,” in Proc. 2013 SIAM International Conference on Data Mining. SIAM, 2013, pp. 387–395.

[14] R. Abebe, J. Kleinberg, D. Parkes, and C. E. Tsourakakis, “Opinion dynamics with varying susceptibility to persuasion,” in Proc. 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2018, pp. 1089–1098.

[15] L. Vassio, F. Fagnani, P. Frasca, and A. E. Ozdaglar, “Message passing optimization of harmonic influence centrality,” IEEE Trans. Control of Network Systems, vol. 1, no. 1, pp. 109–120, 2014.

[16] Y. Yildiz, A. Ozdaglar, D. Acemoglu, A. Saberi, and A. Scaglione, “Binary opinion dynamics with stubborn agents,” ACM Trans. Economics and Computation, vol. 1, no. 4, p. 19, 2013.

[17] V. S. Mai and E. H. Abed, “Optimizing leader influence in networks through selection of direct followers,” IEEE Trans. Autom. Control, 2018.

[18] M. Mesbahi and M. Egerstedt, Graph theoretic methods in multiagent networks. Princeton University Press, 2010.

[19] E. Mackin and S. Patterson, “Maximizing diversity of opinion in social networks,” arXiv preprint arXiv:1811.03706, 2018.