A Proof of Section 4.1

A.1 Only one node in the environment

Recall that there is one node with a fixed opinion $p \in [-1, 1]$ in the environment. The opinion of the agent is updated as in Equation (4),

Lemma 3. *If $w + \beta p y(t) + 1 < 0$, the opinion of the agent stays at $\mathrm{sgn}(y(t))$ for all $t' > t$.*

Proof. As shown in Equation (4) that when $w+\beta p y(t)+1 < 0$, $y(t+1) = \mathrm{sgn}(y(t))$. $w + \beta p y(t) + 1 < 0$ is equivalent to $\beta p y(t) < -w - 1 < 0$. Knowing that $|y(t+1)| = 1 \geq |y(t)|$,

$$\beta p y(t+1) < -w - 1$$

Therefore, $y(t') = \mathrm{sgn}(y(t+1)) = \mathrm{sgn}(y(t))$ for all $t' > t$. □

Lemma 4. *If $w+\beta p y(t)+1 > 0$, there exist two fixed points where $y(t+1) = y(t)$: p and $-\frac{1}{\beta p}$. p is attracting while $-\frac{1}{\beta p}$ is repelling.*

Proof. The converged opinion y of the agent should satisfy

$$f(y) = \frac{wy + \beta p^2 y + p}{w + \beta p y + 1}$$

$$f(y) - y = \frac{-\beta p y^2 + (\beta p^2 - 1)y + p}{w + \beta p y + 1} = \frac{u(y)}{v(y)} = 0 \quad (5)$$

where

$$u(y) = -\beta p y^2 + (\beta p^2 - 1)y + p$$
$$v(y) = \beta p y + w + 1$$

By solving $u(y) = 0$, which is equivalent to $f(y) - y = 0$ since $u(y) > 0$, the two fixed points of $f(y)$ are: p and $-\frac{1}{\beta p}$.

Next, we prove that p is attracting and $-\frac{1}{\beta p}$ is repelling.

$$f'(y) = \frac{w(w + \beta p^2)}{(w + \beta p y + 1)^2} \geq 0$$

$|f'(y)| = f'(y)$, then $f'(p) = \frac{w(w+\beta p^2)}{(w+\beta p^2+1)^2} < 1$, thus attracting; while $f'(y - \frac{1}{\beta p}) = \frac{w(w+\beta p^2)}{w^2} > 1$, thus repelling. □

Lemma 5. *If $w + \beta p y(t) + 1 > 0$ and $py(t) \geq 0$, $y = p$.*

Proof. If $p = 0$, $y(t+1) = \frac{w}{w+1} y(t)$, as the iteration goes, $\lim_{t\to\infty} y(t) = 0$;

If $py(t) > 0$, e.g., they are both positive

- when $0 < y(t) < p$, $y(t+1) - y(t) = \frac{u(y(t))}{v(y(t))} > 0$, thus $y(t+1) > y(t)$, the agent's opinion increases until it reaches p;
- when $p < y(t) < 1$, $y(t+1) - y(t) < 0$, the agent's opinion decreases to p.

□

Lemma 6. *If $w + \beta p y(t) + 1 > 0$ and $py(t) < 0$,*

1. If $\left|\frac{1}{\beta p}\right| > 1$, $\lim_{t\to\infty} y(t) = y^e$.

2. If $\left|\frac{1}{\beta p}\right| \leq 1$,

(a) If $|y(t)| < \left|\frac{1}{\beta p}\right|$, $y = p$.

(b) If $y(t) = -\frac{1}{\beta p}$, $y(t') = -\frac{1}{\beta p}$ for all $t' \geq t$.

(c) If $\left|\frac{1}{\beta p}\right| < |y(t)| \leq 1$, $y = \mathrm{sgn}(y(t))$.

Proof. Assume $y(t) \in (0, 1]$ and $p \in (-1, 0)$,

- if $\left|\frac{1}{\beta p}\right| > 1$, all $y(t) \in (0, 1] < -\frac{1}{\beta p}$, $y(t)$ is attracted to p as the updating goes;
- if $\left|\frac{1}{\beta p}\right| = 1$, $y(t)$ is repelled by the extreme point and goes to the attracting one unless it starts with $-\frac{1}{\beta p}$ at time t;
- if $\left|\frac{1}{\beta p}\right| < 1$, when $0 < y(t) < -\frac{1}{\beta p}$, $y(t+1) - y(t) = \frac{u(y(t))}{v(y(t))} < 0$, $y(t+1) < y(t)$, the agent's opinion decreases to p; when $y(t) = -\frac{1}{\beta p}$, $y(t)$ stays there; when $y(t) > -\frac{1}{\beta p}$, $y(t+1) > y(t)$, the agent's opinion increases to the extreme value on its side.

□

A.2 A group of nodes in the environment

Assume there is a set of m neighbour with different fixed opinions, $\mathbf{p} = (p_1, p_2, ..., p_m)$, $m > 1$. We denote

- $q = \sum_j p_j^2$ the sum of the squares of the fixed opinions.
- $s = \sum_j p_j$ the sum of the fixed opinions.
- $m = \sum_j 1$ the number of nodes in the environment.

Lemma 7. *$mq - s^2 \geq 0$, which is $m \sum_j p_j^2 \geq (\sum_j p_j)^2$.*

Proof.

$$m \sum_j p_j^2 - (\sum_j x_j)^2 = \frac{1}{2} \sum_i \sum_j (p_i - p_j)^2 \geq 0$$

□

The agent's opinion is updated by

$$y(t+1) = \begin{cases} \mathrm{sgn}(y(t)) & \text{if } w + \beta s y(t) + m \leq 0, \\ \frac{wy(t)+\beta q y(t)+s}{w+\beta s y(t)+m} & \text{otherwise.} \end{cases} \quad (6)$$

Lemma 8. *If $w+\beta s y(t)+m > 0$, there exist two fixed points where $y(t+1) = y(t)$:*

$$y^a = \frac{\beta q - m + \sqrt{\Delta}}{2\beta s} \qquad y^r = \frac{\beta q - m - \sqrt{\Delta}}{2\beta s}$$

where $\Delta = (\beta q - m)^2 + 4\beta s^2$. y^a is attracting while y^r is repelling.

Proof. The function is $f(y) = \frac{wy+\beta q y+s}{w+\beta s y+m}$. The two fixed points satisfy $f(y) = y$. $|f'(y)| = f'(y)$ since

$$f'(y) = \frac{(w + \beta q)(w + m) - \beta s^2}{(\beta s y + w + m)^2} = \frac{w(w + m) + \beta q w + \beta(qm - s}{(\beta s y + w + m)^2}$$

For $y^a = \frac{\beta q - m + \sqrt{\Delta}}{2\beta s}$, $f'(y^a) < 1$ because

$$f'(y^a) - 1 = -\frac{1}{2}\frac{(m-\beta q)^2 + 4\beta s^2 + (2w+m+\beta q)\sqrt{\Delta}}{(\beta s y^a + w + m)^2} < 0$$

For $y^r = \frac{\beta q - m - \sqrt{\Delta}}{2\beta s}$, $f'(y^r) > 1$ because

$$f'(y^r) - 1 = -\frac{1}{2}\frac{(m-\beta q)^2 + 4\beta s^2 - (2w+m+\beta q)\sqrt{\Delta}}{(\beta s y^r + w + m)^2} = -\frac{1}{2}\frac{A}{B}$$

$\frac{A}{B} < 0$ since $B > 0$ and it can be proved as below that $A < 0$.

$$\begin{aligned} &\left[(m-\beta q)^2 + 4\beta s^2\right]^2 - \left[(2w+m+\beta q)\sqrt{\Delta}\right]^2 \\ =&4\left[(m-\beta q)^2 + 4\beta s^2\right]\left[\beta(s^2 - qm) - w(m+w+\beta q)\right] < 0 \end{aligned}$$

Therefore, y^a is attracting and y^r is repelling. □

B Proof of Theorem 2

Recall that $y_i(t) \in (-1, 0) \cup (0, 1)$. Given any opinion vector $\mathbf{y}(0)$ of a given connected network $G = (V, E)$, the opinions can be divided into two groups V_1 and V_2 at any time t: a) $\forall i \in V_1$, $y_i(t) > 0$; b)$\forall i \in V_2$, $y_i(t) < 0$, and $V = V_1 \cup V_2$. Denote $n_i^s(t)$ the number of node i's neighbors node that are in the **s**ame group with i at time t, and $n_i^d(t)$ the number of neighbors in the **d**ifferent group. Specifically, they are denoted as

$$\begin{aligned} n_i^s(t) =&|N(i)^s|, N(i)^s = \{j | j \in N(i), \text{ and } y_i(t)y_j(t) > 0\} \\ n_i^d(t) =&|N(i)^d|, N(i)^d = \{k | k \in N(i), \text{ and } y_i(t)y_k(t) < 0\} \end{aligned}$$

Lemma 9. *For node $i \in V$ fix $\beta_i = \beta > 0$, if $\beta > \frac{1}{[\min(|\mathbf{y}(0)|)]^2}$, $\lim_{t\to\infty} |y_i(t)| = 1$.*

Proof. For node $i \in V$, the opinion is updated as in Equation (3). If $\gamma = 1 + \sum_{j \in N(i)} w_{ij} \le 0$, $y_i(t+1)$ reaches the extreme value in one iteration due to strong backfire effect.

While when $\gamma > 0$, for any $t > 0$, $y_i(t+1)$ is updated as

$$y_i(t)\frac{1 + \sum_{j\in N(i)^s} w_{ij}\frac{y_j(t)}{y_i(t)} + \sum_{k \in N(i)^d} w_{ik}\frac{y_k(t)}{y_i(t)}}{1 + \sum_{j \in N(i)^s} w_{ij} + \sum_{k\in N(i)^d} w_{ik}} = y_i(t)\frac{C}{D} \tag{7}$$

When $\beta > \frac{1}{[\min(|\mathbf{y}(t)|)]^2}$, for all $k \in N(i)^d$, $w_{ik} = \beta y_i(t) y_k(t) + 1 < 0$. The sums in Equation (7) satisfy: $\sum_{j\in N(i)^s} w_{ij}\frac{y_j(t)}{y_i(t)}$, $\sum_{j\in N(i)^s} w_{ij}$, $\sum_{k\in N(i)^d} w_{ik}\frac{y_k(t)}{y_i(t)} > 0$, and $\sum_{k \in N(i)^d} w_{ik} < 0$.

Now we focus on the node that has the most moderate opinion, namely the node with absolute value of opinion $\min|\mathbf{y}(t)|$ at each time step, starting from time 0. Knowing $C, D > 0$,

$$C - D = \sum_{j \in N(i)^s} w_{ij}(\frac{y_j(t)}{y_i(t)} - 1) + \sum_{k\in N(i)^d} w_{ik}(\frac{y_k(t)}{y_i(t)} - 1) \tag{8}$$

Since $y_i(t)$ has the smallest absolute opinion value, for any $j \in N(i)^s$, $\frac{y_j(t)}{y_i(t)} \ge 1$, thus $C > D$, $\frac{C}{D} > 1$, and $|y_i(t+1)| > |y_i(t)|$.

After every iteration from time t to $t+1$, the opinion of the most moderate node becomes more extreme, until it reaches the absolute value of 1, thus for any $i \in V$, $\lim_{t\to\infty} |y_i(t)| = 1$. □

Lemma 10. *For node $i \in V$, if $\beta < \frac{1}{[\max(|\mathbf{y}(0)|)]^2}$, there exists a unique $y^* \in [-\max(|\mathbf{y}(0)|), \max(|\mathbf{y}(0)|)]$ such that $\lim_{t\to\infty} y_i(t) = y^*$ for all $i \in V$.*

Proof. When $\beta < \frac{1}{[\max(|\mathbf{y}(0)|)]^2}$, $\gamma = 1 + \sum_{j \in N(i)} w_{ij} > 0$ because for any $j \in N(i)$, $w_{ij} = \beta y_i(t) y_j(t) + 1 > 0$.

For any $t > 1$, $y_i(t+1)$ is updated as in Equation (7), however, the sums have different values: $\sum_{j\in N(i)^s} w_{ij}\frac{y_j(t)}{y_i(t)}$, $\sum_{j\in N(i)^s} w_{ij}$, $\sum_{k\in N(i)^d} w_{ik} > 0$, and $\sum_{k\in N(i)^d} w_{ik}\frac{y_k(t)}{y_i(t)} < 0$.

Then we focus on the most opinionated node, which means the node has the largest absolution value of its opinion $\max|\mathbf{y}(t)|$, starting from time 0. Knowing $D > 0$,

- when $C > 0$, $C - D$ is shown in Equation (8). With i being the most opinionated node, $\frac{y_j(t)}{y_i(t)} \le 1$ for all $j \in N(i)^s$; $\frac{y_k(t)}{y_i(t)} < 0$ for all $k \in N(i)^d$. Therefore, $C < D$, $0 < \frac{C}{D} < 1$ and $|y_i(t+1)| < |y_i(t)|$.
- when $C = 0$, $y_i(t+1) = 0$.
- when $C < 0$, $-C - D$ is shown in Equation (9). As $-1 \le \frac{y_k(t)}{y_i(t)} \le 0$ for $k \in N(i)^d$, $-C - D < 0$, $0 < \left|\frac{C}{D}\right| < 1$, thus $|y_i(t+1)| < |y_i(t)|$.

$$-C - D = -2 - \sum_{j\in N(i)^s} w_{ij}(\frac{y_j(t)}{y_i(t)} + 1) - \sum_{k \in N(i)^d} w_{ik}(\frac{y_k(t)}{y_i(t)} + 1) \tag{9}$$

At every time step, the most opinionated node get moderated until they reach consensus - there is no such node and the updating process stops because consensus is reached.

□

Lemma 11. *For node $i \in V_1$, $y_i(0) = y_0$, where $0 < y_0 < 1$; $\forall i \in V_2$, $y_i(0) = -y_0$. If $\beta = \frac{1}{y_0^2}$, $y_i(t) = y_i(0)$ for all $t \ge 0$.*

Proof. When $\beta = \frac{1}{y_0^2}$, $w_{ij} = \frac{1}{y_0^2} y_i(t) y_j(t)$. At time 1,

$$y_i(1) = \frac{y_i(0) + 2n_i^s(0) y_i(0)}{1 + 2n_i^s(0)} = y_i(0)$$

For any $t \ge 1$,

$$y_i(t+1) = \frac{y_i(t) + 2n_i^s(t) y_i(t)}{1 + 2n_i^s(t)} = y_i(t) = y_i(0)$$

. □