Networks, Cascades, and Influence

Seminar on Markets, Algorithms, Incentives, and Networks - WS21/22

Leon Zamel - **Technical University of Munich**

How do people influence each other? Why might an opinion spread through a network? And how do these answers change under different payoff models and network structures? We will give mathematical models to answer these questions involving social structures.

1 Preliminaries[1]

A *network* is modeled as an *undirected graph* G:

- *Vertices* represent agents/players: $N = \{1, ..., n\}$.
- *Edges* represent a relationship between two agents.

Players can take *actions*:

- The *action space* $A = A_1 \times ... \times A_n$, represents available actions for each agent. For simplicity we only cover binary action games $A_i = \{0, 1\}, \forall i \in N$.
- An *action profile* $a = (a_1, ..., a_n) \in A$, denotes actions taken by each agent i from their set of actions A_i.
- $u_i(a)$ denotes agent i's *utility/payoff* in action profile a.

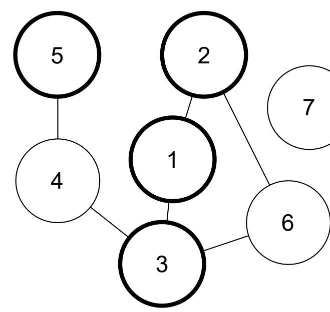

Fig. 1.1: A network with $n = 7$. Agents playing action 1 are shown bold.

Agents only interact with and are influenced by their *neighborhood*:

- The *neighborhood* $N_i \subseteq N$ denotes the players which i has a relationship to.
- An agent's utility only depends on their own action and the number of 0 and 1 actions taken by agents in their neighborhood.
- The *degree* $d_i = |N_i|$ represents the number of neighbors i has.
- $z_i(b, a_{-i}; G) \in \{0, ..., n\}$ denotes how many neighbors of player i play action b, and similarly $f_i(b, a_{-i}; G) \in [0, 1]$ denotes the fraction of agents in i's neighborhood playing action b.

A simple example is given in Figure 1.1. For agent 3, one of their three neighbors plays action 1.

In general, there are two main settings imaginable for how the utility behaves. With an increase of neighbors taking action 1, the payoff of playing 1 vs 0 can either (weakly) *increase* or *decrease*, we will examine both cases in the following.

[1] All definitions are based on [1, 2] and are marked explicitly if taken verbatim from there.

2 Public Goods Game

From an economic standpoint, a public good can be "used" by more than one person and using it multiple times does not decrease its value. In a setting of networks, this could be something bought by one person, but is then used by people in their neighborhood as well. E.g.: a video streaming service subscription can also be used by friends of the buyer. This motivates our first definitions.

Definition 1. Strategic substitutes [1]: A network game on graph G with binary actions 0 and 1 satisfies strategic substitutes if, for every agent i,

$$u_i(1, a'_{-i}) - u_i(0, a'_{-i}) \leq u_i(1, a_{-i}) - u_i(0, a_{-i}),$$

for all a'_{-i} and a_{-i} such that $f_i(1, a'_{-i}; G) \geq f_i(1, a_{-i}; G)$.

Intuitively, the more neighbors already play action 1, the less beneficial it is to play action 1 compared to action 0.

Definition 2. Public goods game (PGG) [1]: The public goods game on graph G is a network game parameterized by cost c, with $0 < c < 1$. Agent i's utility is

$$u_i(a_i, a_{-i}) = \begin{cases} 1 - c & \text{if } a_i = 1, \\ 1 & \text{if } a_i = 0 \text{ and } f_i(1, a_{-i}; G) > 0, \text{ and} \\ 0 & \text{if } a_i = 0 \text{ and } f_i(1, a_{-i}; G) = 0. \end{cases}$$

The public goods game satisfies the strategic substitutes property, as playing action 1 compared to 0 will yield a difference of $1 - c$ if no neighbor plays 1 and a difference of $-c$ otherwise.

Theorem 1. Per [1]: *An action Profile is a pure-strategy Nash equilibrium in the public goods game on graph G if and only if $X_1 = \{i : a_i = 1\}$ is a maximal independent set on graph G.*

A maximal independent set (MIS) is a set of vertices on a graph G where no two vertices of said set have a common edge (they are independent) and no vertex can be added to the set without violating the independence property (the set is maximal).

The theorem can thus be interpreted as the best strategy being to play action 1 and therefore "paying" for the good only if no neighbor plays 1. Because if a neighbor already pays for the good, one can simply "freeride". Given this equivalence, finding a Nash equilibrium is as simple as finding a MIS, which can be done with Algorithm 1. An example of a NE in a PGG is seen in Figure 2.1.

Algorithm 1 Find a NE in a PGG

Input: Graph G defined on agents N
Variables: R: the set of unassigned agents; a_i: the action of agent i

$R := N$, $a_i := 0$ for all i
while $R \neq \emptyset$ **do**
  $i :=$ agent selected randomly from R
  $a_i := 1$
  $R := R - \{i \cup N_i\}$
end while
return $(a_1, ..., a_n)$

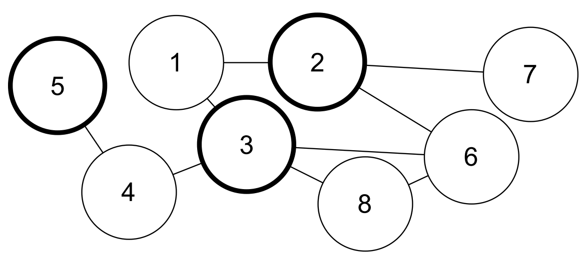

Fig. 2.1: A network with vertices of a MIS shown bold. The MIS corresponds to a NE of a PGG if the agents of that set play 1.

3 Coordination Game

Contrary to PGGs, in coordination games one wants their neighborhood to play the same action as oneself. Common examples are the adoption of technology, language, etc.

Definition 3. Strategic complements [1]: A network game on graph G with binary actions 0 and 1 satisfies strategic complements if, for every agent i,

$$u_i(1, a'_{-i}) - u_i(0, a'_{-i}) \geq u_i(1, a_{-i}) - u_i(0, a_{-i}),$$

for all a'_{-i} and a_{-i} such that $f_i(1, a'_{-i}; G) \geq f_i(1, a_{-i}; G)$

The key (and only) difference to strategic substitutes is the direction of the weak inequality term in the first part of the definition. The more neighbors play action 1 vs action 0, the more beneficial it becomes to play action 1 oneself.

Definition 4. Coordination game [1]: The coordination game on graph G is a network game parameterized by payoff q, with $0 < q < 1$. Agent i's utility is

$$u_i(a_i, a_{-i}) = \begin{cases} (1-q) \cdot z_i(1, a_{-i}; G) & \text{if } a_i = 1, \text{ and} \\ q \cdot z_i(0, a_{-i}; G) & \text{if } a_i = 0. \end{cases}$$

Where q is the utility for each neighbor playing action 0 if player i also plays action 0, and vice versa for $(1 - q)$ and action 1.

This game satisfies the strategic complements property, as the difference in utility for player i playing action 1 vs 0 is

$$(1-q) \cdot z_i(1, a_{-i}; G) - q \cdot z_i(0, a_{-i}; G) = z_i(1, a_{-i}; G) - q \cdot d_i,$$

which strictly increases with the number of neighbors who play action 1.

We see that the term will be (weakly) positive for $z_i(1, a_{-i}; G) \geq q \cdot d_i \iff f_i(1, a_{-i}; G) \geq q$, in words, once the fraction of players in the neighborhood of agent i playing 1 is greater than q. Action 1 is therefore a best response if at least fraction q of the agent's neighborhood plays 1 and action 0 is a best response otherwise. Knowing when an action is a best response, we are interested in Nash equilibria. The following property of a set of players is very useful for this.

Definition 5. Cohesiveness: On an undirected graph G the cohesiveness of a set of players X is

$$coh_G(X) = \min_{i \in X} \frac{|N_i \cap X|}{|N_i|}$$

Intuitively, the cohesiveness gives a lower bound on how "well connected" a set of players is, by measuring for each agent of the set the fraction of neighbors who are also in the set, and taking the minimum of those values.

Theorem 2. Per [1]: *An action profile* a *is a pure-strategy Nash equilibrium in the coordination game with parameter q on graph G if and only if the sets $X_0 = \{i : a_i = 0\}$ and $X_1 = \{i : a_i = 1\}$ satisfy:*

1. $coh_G(X_0) \geq 1 - q$, *and*
2. $coh_G(X_1) \geq q$.

In words, in a Nash equilibrium of a coordination game, every player must have enough neighbors who play the same action as themselves such that their action is a best response.

4 Hard-Threshold Cascades

From the coordination games we see how agents might influence each other once a certain *threshold* of neighbors playing a certain action is reached. If this causes other agents to change their action, this explains how an action/opinion might *cascade* through a network. We therefore introduce a new model in which we will study 0 to 1 cascades happening over discrete *time periods* $t \in \{0, 1, 2, ...\}$.

- A *full cascade* occurs if at the end of the cascade all players have adopted action 1.
- A *partial cascade* occurs if at the end of the cascade some players switched from action 0 to 1 but none from 1 to 0.

Definition 6. Hard-threshold cascade model: On an undirected graph G with threshold parameter $0 < q < 1$ and a seed set S of agents, let time period $t = 0$ be the initialization period. Then an agent selects an action in period $t \geq 0$ according to the rules:

- Each seed agent $i \in S$ plays action 1.
- All other agents $i \in N \setminus S$ play: 0 in $t = 0$ and for period $t > 1$ play action 1 if fraction q or more agents in N_i play 1 in period $t - 1$ (tie breaks in favor of action 1), and 0 otherwise.

The hard-threshold cascade terminates in the first period where no agent has changed their action.

Theorem 3. Per [1]: *The set of agents who play* 1 *is monotonically non-decreasing in each period of the hard-threshold cascade model, and the process terminates.*

We relate this model and the coordination game by introducing the *restricted coordination game*, which adds the constraint to the coordination game that agents in the seed set can only play action 1. For a Nash equilibrium we only expect of non-seed agents to play a best response.

Theorem 4. Per [1]: *A hard-threshold cascade with threshold q and seed set S terminates with a Nash equilibrium of the restricted coordination game with payoff parameter q.*

Knowing this, we are interested in determining when a seed set might create a full cascade. Once more, the cohesion gives an important insight.

Theorem 5. Per [1]: *There is a full cascade from seed set S in the hard-threshold cascade model on graph G and with threshold q if and only if there is no set $T \subseteq N \setminus S$ for which $coh_G(T) > 1 - q$.*

Therefore, we can view a set T with such a property as *blocking* a full cascade. An example cascade is shown in Figure 4.1. The seed agents 3 and 6 cause a partial cascade.

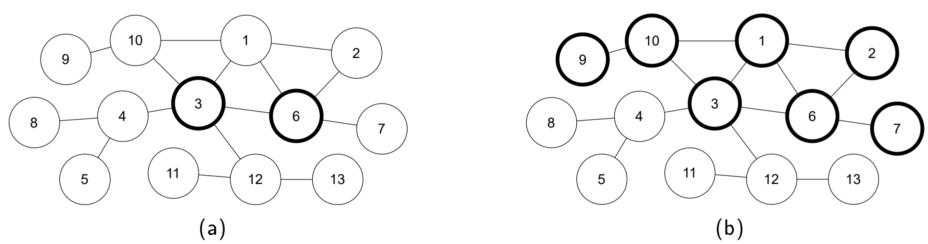

Fig. 4.1: An example of the hard-threshold cascade model with $q = 2/5$ and $S = \{3, 6\}$. Agents playing 1 shown bold. (a) The initialization period. (b) The final actions after the cascade.

5 Independent Cascades

In contrast to the hard-threshold cascade model before, we could imagine a setting where agents switch from 0 to 1 with some probability whenever one of their neighbors switches from 0 to 1. This makes sense when modeling word-of-mouth effects or the spread of diseases.

Definition 7. Independent-cascade model: On a **weighted and directed** graph G we define the independent-cascade model. There is a set of seed agents S. An agent is *activated* in period t if they adopt action 1 for the first time in period t, or they are a seed agent and $t = 0$, the *initialization period*. The weight p_{ij}, with $0 < p_{ij} < 1$, for a directed edge from i to j denotes the *influence probability* of i on j. Agents select their action in period $t \geq 0$ as follows:

- Each seed agent $i \in S$ plays action 1.
- All other agents $i \in N \setminus S$ play: 0 in $t = 0$ and 1 if they played 1 in period $t - 1$. If player i played action 0 in $t - 1$, consider each neighbor $j \in N_i$ who was activated in period $t - 1$ (including seeds if $t = 1$), adopt action 1 with probability p_{ji}.

It is interesting to determine how good a seed set is at causing a cascade (and therefore influencing others). We denote with $L_G(S)$ a random variable corresponding to the set of agents playing action 1 at the end of a cascade from seed set S.

Definition 8. Influence function: The influence function gives the expected independent-cascade size (including seeds):

$$h_G(S) = \mathbb{E}\left[\big|L_G(S)\big|\right].$$

A logical next step is trying to optimize the initial seed set to maximize this influence.

Definition 9. Influence-maximization problem (INFLUENCE): The influence-maximization problem (INFLUENCE) for a directed and weighted graph G and cardinality $k \geq 1$ is defined as solving

$$\max_{S \subseteq N : |S| \leq k} h_G(S).$$

It is, however, difficult in practice to optimize this. Just evaluating the influence function is already non-trivial [3]. We can formalize the *hardness* of this problem if we look at the decision version of INFLUENCE, i.e., answering the question if there is a seed set S with $h_G(S) \geq w$, for some value w and cardinality k.

Theorem 6. Per [1]: INFLUENCE *in the independent-cascade model is* ***NP****-hard.*

This can be shown via a reduction of the set-cover problem.

While this strongly suggests that there is no "fast" algorithm for finding a solution, there are greedy algorithms which can find good approximate solutions. One such algorithm always adds the agent i to the seed set S which will maximize $h_G(S \cup \{i\})$ for that iteration. This is done k times.

References

[1] David C. Parkes and Sven Seuken, Economics and Computation (Draft: November 19, 2018)

[2] Matthew O. Jackson and Yves Zenou, Games on Networks (2014)

[3] Christian Borgs and Michael Brautbar and Jennifer Chayes and Brendan Lucier, Maximizing Social Influence in Nearly Optimal Time (2016)