Optimal Seeding in Large-Scale Super-Modular Network Games

Original
Optimal Seeding in Large-Scale Super-Modular Network Games / Messina, Sebastiano; Cianfanelli, Leonardo; Como, Giacomo; Fagnani, Fabio. - In: IEEE CONTROL SYSTEMS LETTERS. - ISSN 2475-1456. - 8:(2024), pp. 1811-1816. [10.1109/lcsys.2024.3418308]

Availability:
This version is available at: 11583/2990423 since: 2024-07-06T12:31:50Z

Publisher:
IEEE

Published
DOI:10.1109/lcsys.2024.3418308

(Article begins on next page)

Optimal Seeding in Large-Scale Super-Modular Network Games

Sebastiano Messina, Leonardo Cianfanelli, Giacomo Como, and Fabio Fagnani

***Abstract*— We study optimal seeding problems for binary super-modular network games. The system planner's objective is to design a minimal cost seeding guaranteeing that at least a predefined fraction of the players adopt a certain action in every Nash equilibrium. Since the problem is known to be NP-hard and its exact solution would require full knowledge of the network structure, we focus on approximate solutions for large-scale networks with given statistics. In particular, we build on a local mean-field approximation of the linear threshold dynamics that is known to hold true on large-scale locally tree-like random networks. We first reduce the optimal intervention design problem to a linear program with an infinite set of constraints. We then show how to approximate the solution of the latter by standard linear programs with finitely many constraints. Our solutions are then numerically validated.**

I. Introduction

Designing interventions by a central planner in order to modify the outcomes of a network game and steer them towards a socially desirable objective is a fundamental problem in many multi-agent systems. Applications in socio-technical systems are countless, ranging from pricing and toll design in transportation and energy networks to viral marketing in social networks. The optimal intervention design problem for network games is known to be challenging since an intervention on a single individual or on a group of them has direct and indirect effects on all the others. Such spill-over effects depend both on the geometry of the network and on the type of influence mechanisms that individuals' actions have on their neighbors' utilities (e.g., strategic complements vs strategic substitutes) [1]–[3]. Especially over the past two decades, a large body of literature has in particular highlighted the role of network centrality measures in order to determine the network nodes that the intervention should target in order to optimize its effect [4]–[6].

Particularly when dealing with large-scale network games, a central planner faces two key challenges in designing interventions. First, optimal intervention strategies do not scale well with the network size: e.g., many formulations of optimal seeding problems are NP-hard, so that one is led to seeking approximating algorithms for their sub-optimal solution [7]–[10]. Second, full information on the network structure is often not available to the planner as it is either too expansive to collect or severely constrained by proprietary and privacy issues. In these cases, the central planner might rely on a (random) network model that matches the available information (e.g., some statistics) on the network and use this to design the intervention, see, e.g., [11].

In this paper, we focus on optimal seeding in large-scale super-modular network games with binary action sets [9] and on the related linear-threshold dynamics [12], that have received a large amount of attention, since the seminal work [7]. Specifically, we study the problem of finding a minimal cost seeding that guarantees that at least a predefined fraction of the players adopt a certain action in every Nash equilibrium. Building on the local mean-field approximation of the linear threshold dynamics on the configuration model random graph [13], we set up an optimization problem that depends only on the network statistics. We show that such optimization problem turns out to be a linear program with an infinite set of constraints. We then show how to approximate the solution by standard linear programs with finitely many constraints. Our solutions are then validated on both random and deterministic networks.

The paper is organized as follows. Section II formulates the optimal seeding problem. In Section III, we formulate an approximated problem for large-scale networks. Section IV illustrates how to reduce the optimization problem to a linear problem with finite number of constraints. Section V presents some numerical experiments.

Notation The all-one and all-zero vectors are denoted by $\mathbf{1}$ and $\mathbf{0}$, respectively. Inequalities between vectors are meant to hold entry-wise.

II. Optimal seeding problem

We model networks as finite directed multi-graphs $\mathcal{N} = (\mathcal{V}, \mathcal{E}, \theta, \lambda)$, where $\mathcal{V}$ is the node set, $\mathcal{E}$ is the set of directed links, and $\theta : \mathcal{E} \to \mathcal{V}$ and $\lambda : \mathcal{E} \to \mathcal{V}$ are the maps associating to each link its tail and head node, respectively. Let $n = |\mathcal{V}|$ be the network order, and A in $\mathbb{R}^{\mathcal{V}\times\mathcal{V}}$ be the adjacency matrix, with entries $A_{ij} = |\{e \in \mathcal{E} : \theta(e) = i, \lambda(e) = j\}|$. Let $\kappa = A\mathbf{1}$ and $\delta = A'\mathbf{1}$ denote the out- and in-degree vectors, respectively. We assume that $\mathcal{N}$ contains no self-loops, i.e., $\theta(e) \neq \lambda(e)$ for every e in $\mathcal{E}$.

Given a network $\mathcal{N} = (\mathcal{V}, \mathcal{E}, \theta, \lambda)$, we consider binary-action semi-anonymous (BASA) games $\mathcal{G}$ on $\mathcal{N}$ [14]–[16], i.e., strategic games with player set $\mathcal{V}$, whereby every player i in $\mathcal{V}$ has action set $\mathcal{A} = \{0, 1\}$ and utility function

$$u_i(x_i, x_{-i}) = f_i\left(x_i, \sum_j A_{ij}x_j\right), \qquad \forall x \in \mathcal{X}, \quad (1)$$

The authors are with the Department of Mathematical Sciences, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129, Torino, Italy. Email: {sebastiano.messina, leonardo.cianfanelli, giacomo.como, fabio.fagnani}@polito.it. Giacomo Como is also with the Department of Automatic Control, Lund University, Lund, Sweden.

This study was carried out within the FAIR - Future Artificial Intelligence Research and received funding from the European Union Next-GenerationEU (PIANO NAZIONALE DI RIPRESA E RESILIENZA (PNRR) – MISSIONE 4 COMPONENTE 2, INVESTIMENTO 1.3 – D.D. 1555 11/10/2022, PE00000013). This manuscript reflects only the authors' views and opinions, neither the European Union nor the European Commission can be considered responsible for them.

where $f_i : \mathcal{A} \times \mathbb{Z}_+ \to \mathbb{R}$ and $\mathcal{X} = \mathcal{A}^{\mathcal{V}}$ is the strategy profile space. We shall further assume that the function

$$g_i : \mathbb{Z} \to \mathbb{R}\,, \qquad g_i(s) = f_i\,(1, s) - f_i\,(0, s)\;, \qquad (2)$$

returning the net utility gain of a player i, when unilaterally switching from action 0 to action 1, is non-decreasing in the number $s = \sum_j A_{ij} x_j$ of her out-neighbors that are playing action 1. For BASA games, this is equivalent to the so-called increasing difference property, hence to their super-modularity [9]. Thus, we refer to BASA games such that $g_i(s)$ defined in (2) is non-decreasing as BASASM games.

We shall consider interventions consisting in the choice of a seeding, i.e., a subset of nodes $\mathcal{S} \subseteq \mathcal{V}$ that are forced to play action 1 regardless of the action of the other players. Formally, for $\mathcal{S} \subseteq \mathcal{V}$, we consider the seeded game $\mathcal{G}_{\mathcal{S}}$ where players j in $\mathcal{S}$ have utility function $\overline{u}_j(x_j, x_{-j}) = x_j$ and the rest of the players i in $\mathcal{V} \setminus \mathcal{S}$ have utility function as in (1). A Nash equilibrium for the seeded game $\mathcal{G}_{\mathcal{S}}$ is a strategy profile x^* in $\mathcal{X}$ such that

$$x^*_j = 1,\, \forall j \in \mathcal{S}, \quad u_i(x^*_i, x^*_{-i}) \geq u_i(1-x^*_i, x^*_{-i}),\, \forall i \in \mathcal{V}\backslash\mathcal{S}.$$

The set of Nash equilibria of the seeded game $\mathcal{G}_{\mathcal{S}}$ is denoted by $\mathcal{X}^*_{\mathcal{S}}$. The following result states that the best response correspondence admits a threshold structure, and that a minimal Nash equilibrium of $\mathcal{G}_{\mathcal{S}}$ exists for every $\mathcal{S} \subseteq \mathcal{V}$.

Lemma 1: For a BASASM game $\mathcal{G}$ on a network $\mathcal{N}$, let

$$\rho_i = \begin{cases} -1 & \text{if} \quad g_i(0) > 0 \\ \max\{s \in \{0, \ldots, \kappa_i\} \colon g_i(s) \leq 0\} & \text{if} \quad g_i(0) \leq 0 \end{cases} \qquad (3)$$

for every player i in $\mathcal{V}$. Then:

(i) the utility of a player i in $\mathcal{V}$ sastisfies

$$u_i(1, x_{-i}) > u_i(0, x_{-i})\,, \qquad (4)$$

if and only if

$$\sum\nolimits_j A_{ij} x_j > \rho_i\,; \qquad (5)$$

(ii) for every seeding $\mathcal{S} \subseteq \mathcal{V}$ there exists a *minimal Nash equilibrium* of $\mathcal{G}_{\mathcal{S}}$, i.e., $\underline{x}^{\mathcal{S}}$ in $\mathcal{X}^*_{\mathcal{S}}$ such that $\underline{x}^{\mathcal{S}} \leq x^*$ for every other Nash equilibrium x^* in $\mathcal{X}^*_{\mathcal{S}}$.

Proof: (i) It is immediate from (1) and (2) that (4) is equivalent to $g_i(\sum_j A_{ij} x_j) > 0$. Since g_i is non-decreasing, this is in turn equivalent to (5).

(ii) That $g_i(s)$ is non-decreasing for every i in $\mathcal{V}$ implies super-modularity of the seeded game $\mathcal{G}_{\mathcal{S}}$, for every $\mathcal{S} \subseteq \mathcal{V}$. Existence of a minimal Nash equilibrium $\underline{x}^{\mathcal{S}}$ then follows from standard results for super-modular games [9], [17]. ■

Remark 1: Lemma 1 implies that the minimal Nash equilibrium $\underline{x}^{\mathcal{S}}$ of the seeded game $\mathcal{G}_{\mathcal{S}}$ only depends on the network $\mathcal{N}$, the threshold vector ρ, and the seeding $\mathcal{S}$. Clearly, if $\rho = -\mathbf{1}$, then $\underline{x}^{\emptyset} = \mathbf{1}$. Notice that including a player i in the seeding $\mathcal{S}$ effectively amounts to replacing its threshold ρ_i with a new value $\overline{\rho}_i = -1$ and that $\underline{x}^{\mathcal{V}} = \mathbf{1}$.

We can now formulate our optimal seeding problem. Let $\gamma_i \geq 0$ be the cost associated to including a player i in the seeding $\mathcal{S}$. Without loss of generality (c.f., Remark 1), we assume that $\gamma_i = 0$ for every i such that $\rho_i = -1$. For a given tolerance value ε in $[0, 1]$, we seek a seeding $\mathcal{S}$ of minimal aggregate cost guaranteeing that

$$\frac{1}{n} \sum_{i \in \mathcal{V}} \underline{x}^{\mathcal{S}}_i \geq 1 - \epsilon\,, \qquad (6)$$

i.e., that the minimal Nash equilibrium of the seeded game $\mathcal{G}_{\mathcal{S}}$ has all but at most a fraction ϵ of players playing action 1. Hence, we are interested in following optimization problem

$$\min \sum_{i \in \mathcal{S}} \gamma_i \,:\, \mathcal{S} \subseteq \mathcal{V} \text{ s.t. (6)}\,. \qquad (7)$$

Observe that (7) is always feasible since $x^{\mathcal{V}} = \mathbf{1}$ (c.f. Remark 1), so that (6) is always satisfied by $\mathcal{S} = \mathcal{V}$.

Observe that an instance of the optimal seeding problem (7) is fully specified by the choice of the tolerance value ε and the triple $(\mathcal{N}, \rho, \gamma)$ of the network $\mathcal{N}$, the threshold vector ρ, and the cost vector γ. It will prove convenient to associate to every player i a type ω_i that uniquely determines its in-degree δ_i, out-degree κ_i, threshold ρ_i, and associated cost γ_i. Let Ω be a countable universe of types and let d, k, r, c in $\mathbb{R}^{\Omega}$ be such that

$$d_{\omega_i} = \delta_i\,, \quad k_{\omega_i} = \kappa_i\,, \quad r_{\omega_i} = \rho_i\,, \quad c_{\omega_i} = \gamma_i\,, \quad \forall i \in \mathcal{V}\,.$$

We shall refer to the empirical distribution of types, i.e., the probability distribution p on Ω with

$$p_w = |\{i \in \mathcal{V} : \omega_i = w\}|/n\,, \qquad \forall w \in \Omega\,, \qquad (8)$$

as the *statistics* of the problem.

A. Optimal seeding in network coordination games

In this subsection, we focus on network coordination games [18], [19] showing how they fit in our setting as a special case. Given a network $\mathcal{N}$ with node set $\mathcal{V}$ and adjacency matrix A, and a vector b in $\mathbb{R}^{\mathcal{V}}$, the coordination game on $\mathcal{N}$ with bias b is the strategic game with player set $\mathcal{V}$, binary action set $\mathcal{A} = \{0, 1\}$, and utility function

$$u_i(x_i, x_{-i}) = \sum_{j \in \mathcal{V}} A_{ij}\,(x_i x_j + (1 - x_i)(1 - x_j)) + b_i x_i\,,$$

for every player i in $\mathcal{V}$. Observe that the utility function above can be rewritten in the form (1) with

$$f_i(1, s) = s + b_i\,, \qquad f_i(0, s) = \kappa_i - s\,.$$

Clearly, we have that

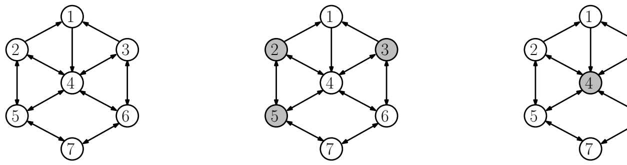
$$g_i(s) = f_i\,(1, s) - f_i\,(0, s) = 2s + b_i - \kappa_i$$

is increasing in s and the threshold values are given by

$$\rho_i = \min\{\max\{-1, \lfloor(\kappa_i - b_i)/2\rfloor\}, \kappa_i\}\,,$$

for every player i in $\mathcal{V}$.

Example 1: Consider the network coordination game $\mathcal{G}$ on the network $\mathcal{N}$ displayed in left of Figure 1(a) with bias $b = 0$. The in- and out-degree vectors are $\delta = (2, 2, 2, 5, 3, 3, 2)$ and $\kappa = (1, 3, 3, 4, 3, 3, 2)$, respectively, while the threshold vector is $\rho = (0, 1, 1, 2, 1, 1, 1)$. Let the cost vector be $\gamma = (2, 1, 1, 4, 1, 1, 2)$ and set the tolerance value ε in $[0, 1/7)$.



(a) (b) (c)

Fig. 1. The network of Example 1.

Then, $\mathcal{S}^* = \{2, 3, 5\}$ (see Figure 1(b)) can be shown to be an optimal seeding with aggregate cost 3 (also $\{2, 3, 6\}$, $\{2, 5, 6\}$, and $\{3, 5, 6\}$ are optimal seedings). On the other hand, $\mathcal{S} = \{4\}$ (see Figure 1(c)) satisfies (6), however its cost is 4, hence it is not an optimal seeding. Finally, notice that nodes can be grouped into five different types by defining $\Omega = \{A, B, C, \dots\}$, $d = (2, 2, 5, 3, 2)$, $k = (1, 3, 4, 3, 2)$, $r = (0, 1, 2, 1, 1)$, and $c = (2, 1, 4, 1, 2)$. The type vector is then $\omega = (A, B, B, C, D, D, E)$ and the problem statistics are $p = (1, 2, 1, 2, 1)/7$.

III. OPTIMAL SEEDING IN LARGE-SCALE NETWORKS

Exactly solving the optimal seeding problem (7) is known to be NP-hard [7], [9]. Our approach consists in first providing an equivalent dynamical representation of the constraint (6), then approximating such dynamics by local mean-field techniques, in the spirit of [13], and finally re-formulating the optimization in this framework. Besides simplifying the analysis, the main advantage of our approach is to allow the design of optimal interventions for families of large-scale networks with given statistics, without requiring full network knowledge.

A. Equivalence with seeded linear threshold dynamics

Given a BASASM game on a network $\mathcal{N}$ and a seeding $\mathcal{S} \subseteq \mathcal{V}$, consider the following discrete-time *seeded linear threshold dynamics* on the strategy profile space $\mathcal{X}$:

$$x(t+1) = \Phi_{\mathcal{S}}(x(t)), \qquad t \geq 0, \tag{9}$$

where $\Phi_{\mathcal{S}} : \mathcal{X} \to \mathcal{X}$ is the map defined by

$$(\Phi_{\mathcal{S}}(x))_i = \begin{cases} 1 & \text{if} \quad i \in \mathcal{S} \text{ or } \sum_j A_{ij} x_j > \rho_i \\ 0 & \text{if} \quad i \notin \mathcal{S} \text{ and } \sum_j A_{ij} x_j \leq \rho_i , \end{cases}$$

and ρ_i are the threshold values defined in (3), for every i in $\mathcal{V}$. Then, we have the following result.

Proposition 1: Given a BASASM game on a network $\mathcal{N}$ and a seeding $\mathcal{S} \subseteq \mathcal{V}$, equation (6) holds true if and only if the seeded linear threshold dynamics (9) is such that

$$\frac{1}{n} \sum_{i \in \mathcal{V}} x_i(t) \geq 1 - \epsilon, \qquad \forall\, t \geq n, \tag{10}$$

for every $x(0)$ in $\mathcal{X}$.

Proof: (6)⇒(10) Since $\Phi_{\mathcal{S}} : \mathcal{X} \to \mathcal{X}$ is monotone non-decreasing and $\Phi_{\mathcal{S}}(\mathbf{0}) \geq \mathbf{0}$, an induction argument shows that, if $x(0) = \mathbf{0}$, then $x(t+1) \geq x(t)$ for every $t \geq 0$. Hence, if $x(0) = \mathbf{0}$, then $x(t)$ converges to a fixed point $x^* = \Phi_{\mathcal{S}}(x^*)$ in at most n steps. Now, notice that every fixed point $x^* = \Phi_{\mathcal{S}}(x^*)$ is a Nash equilibrium of $\mathcal{G}_{\mathcal{S}}$, so that $x^* \geq \underline{x}^{\mathcal{S}}$. Hence, (6) implies that (10) holds true when $x(0) = \mathbf{0}$. Since $\Phi_{\mathcal{S}} : \mathcal{X} \to \mathcal{X}$ is monotone non-decreasing, (10) holds true for all $x(0)$ in $\mathcal{X}$.

(10)⇒(6) Observe that $\Phi_{\mathcal{S}}(x^*) \leq x^*$ for every x^* in $\mathcal{X}^*_{\mathcal{S}}$. Hence, $\Phi_{\mathcal{S}}(\underline{x}^{\mathcal{S}}) \leq \underline{x}^{\mathcal{S}}$. Since $\Phi_{\mathcal{S}} : \mathcal{X} \to \mathcal{X}$ is monotone non-decreasing, an induction argument shows that, if $x(0) = \underline{x}^{\mathcal{S}}$, then $x(t+1) \leq x(t)$ for all $t \geq 0$. Hence, if $x(0) = \underline{x}^{\mathcal{S}}$, then $x(t)$ converges in at most n steps to a fixed point $x^* = \Phi_{\mathcal{S}}(x^*) \leq \underline{x}^{\mathcal{S}}$. Then, (10) with $x(0) = x^*$ implies that $\frac{1}{n} \sum_i \underline{x}^{\mathcal{S}}_i \geq \frac{1}{n} \sum_i x^*_i \geq 1 - \epsilon$, so that (6) holds true. ■

Proposition 1 states that a seeding $\mathcal{S} \subseteq \mathcal{V}$ is such that in the minimal Nash equilibrium of the seeded game $\mathcal{G}_{\mathcal{S}}$ at most a fraction ϵ of players play action 0, if and only if the state $x(t)$ of the seeded linear threshold dynamics (9) has at most a fraction ϵ of entries equal to 0 for every initial state $x(0)$ in $\mathcal{X}$ and time $t \geq n$.

B. Mean-field approximation of linear threshold dynamics

In order to sample uniformly from the set of problems with the same statistics we introduce the following notions. For a probability distribution p on Ω and f in $\mathbb{R}^{\Omega}_{+}$, let

$$\langle p, f \rangle = \sum_{w \in \Omega} p_w f_w \in [0, +\infty] .$$

be the expected value of f with respect to p.

Now, notice that, for a given probability distribution p on Ω and finite set $\mathcal{V}$ of cardinality $n = |\mathcal{V}|$, a type vector ω in $\Omega^{\mathcal{V}}$ whose empirical distribution is p exists if and only if

$$n p_w \in \mathbb{Z}_+ \qquad \forall w \in \Omega . \tag{11}$$

Moreover, for a network $\mathcal{N}$ to exist with node set $\mathcal{V}$ and in- and out-degrees $\delta_i = d_{\omega_i}$ and $\kappa_i = k_{\omega_i}$ for all i in $\mathcal{V}$, it is necessary and sufficient that

$$\langle p, d \rangle = \langle p, k \rangle , \tag{12}$$

$$d_w + k_w \leq n \langle p, d \rangle , \qquad \forall w \in \Omega : \, p_w > 0 . \tag{13}$$

In fact, (12) guarantees that the total out-degree equals the total in-degree, whereas (13) ensures that a wiring exists that does not create any self-loops. We shall refer to the pair (p, n) of a probability distribution p on Ω and a positive integer n as compatible if conditions (11)–(13) are satisfied.

For a consistent pair (p, n), the following definition formalizes the idea of sampling uniformly from the set of problems with statistics p and order n.

Definition 1: Let $\mathcal{V} = \{1, \dots, n\}$ and let ω in $\Omega^{\mathcal{V}}$ have empirical distribution p. Let $l = n \langle p, d \rangle$ and $\mathcal{E} = \{1, \dots, l\}$. Define $\theta : \mathcal{E} \to \mathcal{V}$ by letting $\theta(e)$ in $\mathcal{V}$ be the unique value such that $\sum_{i=1}^{\theta(e)-1} w_i < e \leq \sum_{i=1}^{\theta(e)} w_i$, for all e in $\mathcal{E}$. Let $\lambda : \mathcal{E} \to \mathcal{V}$, $\lambda(e) = \theta(\pi(e))$ for e in $\mathcal{E}$, where π is sampled uniformly from the set of permutations of $\mathcal{E}$ such that

$$\theta(e) \neq \theta(\pi(e)) , \qquad \forall e \in \mathcal{E} . \tag{14}$$

Then, the *(resampled) configuration model* $\mathcal{C}_{n,p}$ is the triple $(\mathcal{N}, \rho, \gamma)$ of the random network $\mathcal{N} = (\mathcal{V}, \mathcal{E}, \theta, \lambda)$, the threshold vector ρ, and the cost vector γ with entries $\rho_i = r_{w_i}$ and $\gamma_i = c_{w_i}$, respectively, for all i in $\mathcal{V}$.

We shall now prove that, with high probability, a random problem sampled from the configuration model $\mathcal{C}_{n,p}$ can be well approximated by a deterministic linear program. Let

$$\varphi_{kr}(z) = \sum_{u=r+1}^{k} \binom{k}{u} z^u (1-z)^{k-u}, \qquad \forall -1 \leq r \leq k\,.$$

Then, define the maps $\psi_p, \phi_p : [0,1] \to [0,1]$ as

$$\psi_p(z) = \sum_{w\in\Omega} p_w \varphi_{k_w r_w}(z)\,,$$

$$\phi_p(z) = \frac{1}{\langle p, d\rangle} \sum_{w\in\Omega} p_w d_w \varphi_{k_w r_w}(z)\,.$$

We can now formalize the following result.

Theorem 1: Let p be a probability distribution on Ω whose moments $\langle p, d\rangle$, $\langle p, k\rangle$, $\langle p, d^2\rangle$, and $\langle p, k^2\rangle$ are all finite. Let (p^n, n) be a sequence of compatible pairs such that

$$p^n_w \overset{n\to+\infty}{\longrightarrow} p_w\,, \quad \forall w\in\Omega\,, \qquad \langle p^n, d\rangle \overset{n\to+\infty}{\longrightarrow} \langle p, d\rangle\,,$$

$$\langle p^n, d^2\rangle \overset{n\to+\infty}{\longrightarrow} \langle p, d^2\rangle\,, \qquad \langle p^n, k^2\rangle \overset{n\to+\infty}{\longrightarrow} \langle p, k^2\rangle\,.$$

If

$$\phi_p(z) > z, \qquad \forall z \in [0, \psi_p^{-1}(1-\epsilon)]\,, \qquad (15)$$

then, for every $\varepsilon > 0$, (6) holds true on the resampled configuration model $\mathcal{C}_{p^n,n}$ with probability converging to 1 as n grows large.

Proof: Let $\nu = \langle p, dk\rangle / \langle p, d\rangle - 1$. Finiteness of the first and second moments of p implies that $0 \leq \nu < +\infty$. An argument analogous to [20, Theorem 7.2] implies that the probability that a uniform random permutation π satisfies (14) converges to $e^{-\nu/2}$ as $n \to +\infty$. Let $\mathcal{S} = \{i : \rho_i = -1\}$ and let $Y(t) = \frac{1}{n}\sum_i x_i(t)$ and $Z(t) = \frac{1}{l}\sum_i (\delta_i x_i(t))$ denote the fraction of action-1 players and the fraction of links pointing to action-1 players in the seeded linear threshold dynamics (9). Define $z(t)$ and $y(t)$ recursively by putting $y(0) = z(0) = 0$ and, for $t \geq 0$,

$$y(t+1) = \psi_p(z(t))\,, \qquad z(t+1) = \phi_p(z(t))\,. \qquad (16)$$

(Figure 2 illustrates the evolution of $z(t)$.) Then, [13, Theorem 1] implies that $Z(t)$ and $Y(t)$ are arbitrarily close to $z(t)$ and $y(t)$, respectively, on all but an asymptotically vanishing fraction of networks $\mathcal{N} = (\mathcal{V}, \mathcal{E}, \theta, \lambda)$, where $\lambda(e) = \theta(\pi(e))$ for a uniform random permutation π of $\mathcal{E}$. Since π satisfies (14) with probability bounded away from 0 asymptotically, this implies that $Z(t)$ and $Y(t)$ are arbitrarily close to $z(t)$ and $y(t)$, respectively, on the configuration model $\mathcal{C}_{n,p^n}$ with probability approaching 1 as n grows large. Observe that (15) and the second equation in (16) imply that there exists a finite time τ such that $z(t) \geq \psi_p^{-1}(1-\epsilon)$ for all $t \geq \tau$. The first equation in (16) then implies that $y(t) \geq 1-\epsilon$ for $t > \tau$. This implies that, with probability converging to 1 as n grows large $\frac{1}{n}\sum_i x_i(t) = Y(t) \geq 1-\epsilon$ for $t > \tau$, i.e., (10) holds true. By Proposition 1, (10) is equivalent to (6), thus completing the proof. ■

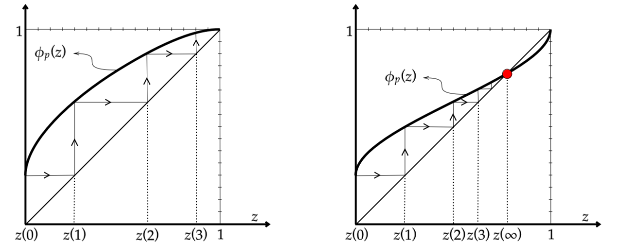


Fig. 2. Plot of the recursion function representing the evolution of the fraction of links pointing to state-1 agents in the mean field approximation.

C. Problem reformulation

In this section we reformulate the optimal seeding problem using the mean-field approximation described in the previous section. The intervention in this framework consists in selecting the fraction of agents of each type that are included in the seeding. This results in a new statistics $\bar{p}$ such that

$$\begin{cases} \bar{p}_w \leq p_w, & \forall w \in \Omega : r_w > -1\,, \\ \bar{p}_w = p_w + \sum_{\substack{w'\in\Omega:\\ d_{w'}=d_w,\\ k_{w'}=k_w,\\ r_{w'}>-1,\\ c_{w'}=c_w}} (p_{w'} - \bar{p}_w) & \forall w\in\Omega : r_w = -1. \end{cases} \qquad (17)$$

Theorem 1 provides us with a sufficient condition for convergence of the threshold dynamics to a configuration with at least a predefined fraction of agents with state 1 for all but a vanishing fraction of networks sampled from $\mathcal{C}(n, \bar{p})$ for large n. Therefore we can reformulate the optimization problem as

$$\inf_{\{\bar{p}\,:\,(17)\}} C(\bar{p}) \quad \text{s.t.} \quad \phi_{\bar{p}}(z) > z\ \forall z \in [0, \psi_{\bar{p}}^{-1}(1-\epsilon)], \quad (18)$$

where $C(\bar{p}) = \langle p - \bar{p}, c\rangle$ denotes the cost for steering the networks statistics from p to $\bar{p}$. Observe that (18) is a linear program, since both the objective function and the constraint are linear in $\bar{p}$. Nonetheless, the problem is very challenging as it is an infinite programming problem, with infinite number of constraints parametrized by z. In the next section we discuss how to solve the problem.

IV. Problem solutions

In this section we propose a numerical method to solve (18) for a tolerance value $\varepsilon > 0$. One of the technical challenges of (18) is that the domain where the constraint $\phi_{\bar{p}}(z) > z$ must be satisfied depends on $\bar{p}$. To avoid this issue, we define

$$\inf_{\{\bar{p}\,:\,(17)\}} C(\bar{p}) \quad \text{s.t.} \quad \phi_{\bar{p}}(z) > z\ \forall z \in [0, 1-\alpha_\epsilon], \qquad (19)$$

where $d_{min} = \min_{i\in\mathcal{V}} \delta_i$ and $\alpha_\epsilon = \epsilon d_{min}/\langle p, d\rangle$. The next result establishes a relation between the two problems.

Proposition 2: Let $\mathcal{P}_\epsilon$ denote the set of admissible $\bar{p}$ for (18) and $\mathcal{P}_{\alpha_\varepsilon}$ denote the set of admissible $\bar{p}$ for (19). Then:

i) $\mathcal{P}_{\alpha_\varepsilon} \subseteq \mathcal{P}_\epsilon$.
ii) $\lim_{\epsilon\to 0} \mathcal{P}_\epsilon \setminus \mathcal{P}_{\alpha_\varepsilon} = \emptyset$.

Proof: Following the same steps of Theorem 1, the constraint in (19) implies that with high probability the dynamics converges to a configuration with at most a fraction α_ε of links pointing to agents with state 0. This implies that the fraction of agents with state 0 is upper bounded by $\alpha_\varepsilon\langle p,d\rangle/d_{min} = \varepsilon$. This in turn proves (c.f. proof of Theorem 1 and [13]) that $\psi_{\bar{p}}(z) \geq 1-\varepsilon$, so that $z \geq \psi_{\bar{p}}^{(-1)}(1-\varepsilon)$ and the constraint in (18) is satisfied. To prove ii), notice that $\psi_{\bar{p}}^{-1}(1) = 1$ for all $\bar{p}$. By continuity of $\psi_{\bar{p}}$ and by definition of α_ε, $\psi_{\bar{p}}^{-1}(1-\epsilon) \overset{\varepsilon\to 0^+}{\longrightarrow} 1$ and $1-\alpha_\epsilon \overset{\varepsilon\to 0^+}{\longrightarrow} 1$, so that the set of admissible interventions for (18) and (19) get arbitrarily close in the limit of small ε. ■

Proposition 2 states that (19) is a restriction of (18), as the set of admissible statistics for the former problem is a subset of the set of admissible statistics for the latter one. Moreover, the two sets converge one to the other one as ϵ vanishes. For this reason, in the rest of the paper we shall focus on (19), which is easier to solve. To handle the infinite number of constraints of the problem, we use the following approach. We discretize the interval $[0, 1-\alpha_\epsilon]$ in $N+1$ equally spaced points $z_i = (1-\alpha_\epsilon)i/N$ with $i = 0, 1, \cdots, N$, and impose that the constraint in (19) is satisfied for each z_i up to a tolerance error Δ_N. The regularity of ϕ_p then guarantees that the solution of the discretized problem is feasible for the original problem, if Δ_N is properly chosen. Given N in $\mathbb{N}$, the discretized problem that we consider is

$$\min_{\{\bar{p}\,:\,(17)\}} C(\bar{p}) \quad \text{s.t.} \quad \phi_{\bar{p}}(z_i) - z_i \geq \Delta_N\,,\ 0 \leq i \leq N, \tag{20}$$

where

$$\Delta_N = \frac{1-\alpha_\epsilon}{2N}\left(\frac{d_{\max}2^{k_{\max}+1}k_{\max}}{\langle p,d\rangle} + 1\right). \tag{21}$$

and $k_{max} = \max_{i\in\mathcal{V}}\kappa_i$, $d_{max} = \max_{i\in\mathcal{V}}\delta_i$. Let $\bar{p}^N$ be the solution of (20) for a given N. The next theorem states that $\bar{p}^N$ is admissible and arbitrarily close to optimal for the original problem (18) as N grows large.

Theorem 2: Let $\bar{p}^*$ be the solution of (18). Then,

i) $\bar{p}^N$ is admissible for (18) for every N.

ii) $C(\bar{p}^N) \overset{N\to+\infty}{\longrightarrow} C(\bar{p}^*)$.

Proof: i) Since $\bar{p}^N$ is solution of (20), it holds $\phi_{\bar{p}^N}(z_i) - z_i \geq \Delta_N$ for every $i = 0, \cdots, N$. Since $z_{i+1} - z_i = (1-\alpha_\epsilon)/N$, by standard properties of the first derivative we establish that for every z in $[0, 1-\alpha_\epsilon]$

$$\phi_{\bar{p}^N}(z) - z > \Delta_N - \frac{(1-\alpha_\epsilon)}{2N}\left|\frac{d}{dz}\left(\phi_{\bar{p}^N}(z) - z\right)\right|. \tag{22}$$

Now, direct algebraic manipulation implies that

$$\left|\frac{d}{dz}(\phi_{\bar{p}^N}(z) - z)\right| \leq \frac{d_{\max}2^{k_{\max}+1}k_{\max}}{\langle p,d\rangle} + 1. \tag{23}$$

From (21) and plugging (23) into (22), it follows that $\phi_{\bar{p}^N}(z) - z > 0$, so that $\bar{p}^N$ is admissible for (19). Proposition 2(i) implies that it is also admissible for (18).

ii) Notice that for all $\bar{p}$ satisfying (17), one can write

$$\phi_{\bar{p}}(z) = \phi_p(z) + \sum_{w\in\Omega} a_w(z)\cdot(p_w - \bar{p}_w), \tag{24}$$

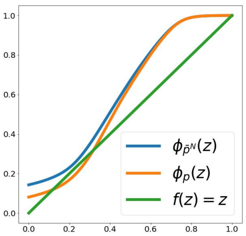


Fig. 3. The recursion functions for the problem presented in Section V.

with $a_w(z) = d_w\left(1-\varphi_{k_w,r_w}(z)\right)/\langle p,d\rangle \geq 0$. Let $\underline{p}$ be the solution of (19) and notice that $\varepsilon > 0$ implies $\alpha_\varepsilon > 0$. This implies the existence of some agents whose threshold is not -1 under statistics $\underline{p}$, namely, there exists a type q in Ω such that $r_q > -1$ and $\underline{p}_q > 0$. Let $\xi_N = \Delta_N/\min_\zeta a_q(\zeta)$ and notice that $\xi_N \overset{N\to+\infty}{\longrightarrow} 0$ since $\Delta_N \overset{N\to+\infty}{\longrightarrow} 0$. Hence, there always exists a sufficiently large N such that $\hat{p}$, defined as

$$\hat{p}_w = \begin{cases} \underline{p}_w - \xi_N, & \text{if } w = q \\ \underline{p}_w + \xi_N & \text{if } (d_w,k_w,c_w) = (d_q,k_q,c_q), r_w = -1 \\ \underline{p}_w & \text{otherwise} \end{cases} \tag{25}$$

satisfies $\hat{p}_q \geq 0$ and is therefore a statistics. Moreover, (24) implies that, for every z in $[0, 1-\alpha_\epsilon]$,

$$\phi_{\hat{p}}(z) = \phi_{\underline{p}}(z) + a_q(z)\cdot\xi_N \geq \phi_{\underline{p}}(z) + \Delta_N > z + \Delta_N, \tag{26}$$

where the last inequality follows from that $\underline{p}$ is admissible for (19) (in fact, it is the solution). Eq. (26) implies that $\hat{p}$ is admissible for (20), whose solution is $\bar{p}^N$. Moreover, $\bar{p}^N$ is admissible for (19), whose solution is $\underline{p}$. Therefore,

$$C(\hat{p}) \geq C(\bar{p}^N) \geq C(\underline{p}). \tag{27}$$

We now prove that $C(\bar{p}^N) \overset{N\to+\infty}{\longrightarrow} C(\underline{p})$. To prove this, we show that $C(\hat{p}) \overset{N\to+\infty}{\longrightarrow} C(\underline{p})$. This follows from (25), which implies $C(\hat{p}) = C(\underline{p}) + c_q\xi_N \overset{N\to+\infty}{\longrightarrow} C(\underline{p})$. That $C(\bar{p}^N) \overset{N\to+\infty}{\longrightarrow} C(\bar{p}^*)$ follows from Proposition 2(ii). ■

Observe that (20) is a linear program with finitely many constraints, therefore much easier to solve than (19). Theorem 2 guarantees that its solution $\bar{p}^N$ is arbitrarily close to optimal as N grows large.

V. Numerical simulations

In this section we present numerical simulations to validate the solutions proposed in Section IV. To this end, we use the topology of the social network Epinions.com[1]. The network contains $n = 26.588$ nodes and 100.120 undirected links, so that $\kappa_i = \delta_i$ for every i in $\mathcal{V}$. It is assumed that every node i has unitary cost $\gamma_i = 1$ and threshold $\rho_i = \lfloor\Theta_i\kappa_i\rfloor - 1$, with Θ_i i.i.d. random variables drawn with uniform probability from $\{0.25, 0.5, 0.75\}$. This defines a triple $(\mathcal{N}, \rho, \gamma)$. We then find the solution $\bar{p}^N$ of the discretized problem (20)

[1]Retrieved from https://networkrepository.com

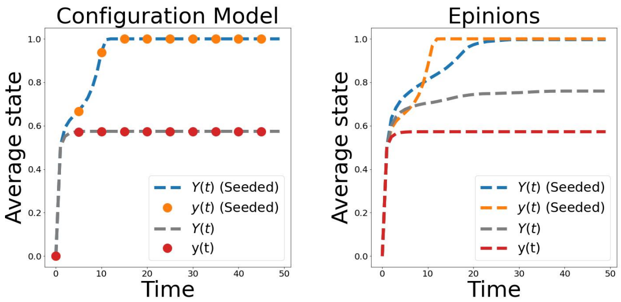


Fig. 4. The TM dynamics on the Epinions.com network (*right*) and on the Configuration model with same statistics (*left*). The plot illustrates the fraction of state-1 adopters $Y(t)$ for the original and the seeded statistics p, $\bar{p}^N$, compared with the outputs $y(t)$ of the recursion (16).

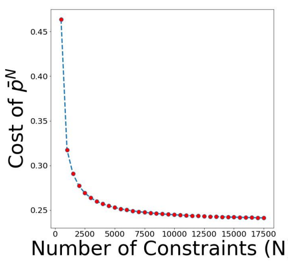


Fig. 5. $C(\bar{p}^N)$ as a function of N for the statistics p' in Section V.

with $N = 10000$ and $\varepsilon = 0.1$ for the statistics p extracted from $(\mathcal{N}, \rho, \gamma)$. Figure 3 illustrates the recursion functions ϕ_p and $\phi_{\bar{p}^N}$ for the original statistics p and for the seeded statistics $\bar{p}^N$. Note that $\phi_{\bar{p}^N}(z) > z$ for all z in $[0, 1]$, so that the dynamics for a game sampled from $\mathcal{C}(n, \bar{p}^N)$ is expected to converge to $x = +\mathbf{1}$ for every initial condition.

To validate this, we generate two instances drawn from $\mathcal{C}(n, p)$ and $\mathcal{C}(n, \bar{p}^N)$ respectively, and simulate the linear threshold dynamics with initial condition $x(0) = \mathbf{0}$. Figure 4 (*left*) illustrates the dynamical behaviour of the average state $Y(t)$ and the one predicted by the recursion (16), denoted $y(t)$, with initial conditions $Y(0) = y(0) = 0$. As expected, since the order of the network is large, the two are very close to each other. Moreover, the average state converges to 1 under statistics $\bar{p}^N$, as predicted by Figure 3. We then apply a seeding consistent with $\bar{p}^N$ to the Epinions network. Despite the lack of theoretical guarantees for this network, the right panel of Figure 4 shows that the dynamics converges to a configuration with almost all ones. This suggests that designing the optimal seeding based on the problem statistics is a valid approach even for real networks that are not generated from a configuration model ensemble.

Finally, to investigate the sensitivity of $\bar{p}^N$ with respect to N, we solve (20) with $\varepsilon = 0.1$ for different values of N for a given statistics $p'_{2,3,0,1} = p'_{2,3,1,1} = p'_{4,3,0,1} = p'_{4,3,1,1} = 1/12$, $p'_{2,3,2,1} = p'_{4,3,2,1} = 1/3$. Figure 5 illustrates the total cost of $\bar{p}^N$ as a function of N. However, notice that N, and therefore the complexity of (20), does not scale with n.

VI. Conclusion

We have studied optimal seeding in large-scale supermodular network games with binary action sets ensuring that all but a small fraction of players chose a certain action in every Nash equilibrium. We have built on local mean-field techniques to approximate the dynamics, formulated a seeding problem in this framework, and proposed an algorithm to transform the resulting infinite programming into a linear program with finitely many constraints. We have then validated the proposed procedure on the Epinions network. Future research includes more general intervention problems where the players' thresholds can be modified.

References

[1] A. Galeotti, S. Goyal, M. Jackson, F. Vega-Redondo, and L. Yariv, "Network games," *The Review of Economic Studies*, vol. 77, no. 1, pp. 218–244, 2010.
[2] A. Galeotti, B. Golub, and S. Goyal, "Targeting interventions in networks," *Econometrica*, vol. 88, no. 6, pp. 2445–2471, 2020.
[3] F. Parise and A. Ozdaglar, "Analysis and interventions in large network games," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 4, pp. 455–486, 2021.
[4] C. Ballester, A. Calvó-Armengol, and Y. Zenou, "Who's who in networks. wanted: The key player," *Econometrica*, vol. 74, no. 5, pp. 1403–1417, 2006.
[5] O. Candogan, K. Bimpikis, and A. Ozdaglar, "Optimal pricing in networks with externalities," *Operations Research*, vol. 60, pp. 883–905, 2012.
[6] L. Damonte, G. Como, and F. Fagnani, "Systemic risk and network intervention," *IFAC PapersOnLine*, vol. 53, no. 2, pp. 2856–2861, 2020.
[7] D. Kempe, J. Kleinberg, and E. Tardos, "Maximizing the spread of influence through a social network," in *Proceedings of the Ninth ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, pp. 137–146, 2003.
[8] D. Kempe, J. Kleinberg, and E. Tardos, "Maximizing the spread of influence through a social network," *Theory of Computing*, vol. 11, no. 4, pp. 105–147, 2015.
[9] G. Como, S. Durand, and F. Fagnani, "Optimal targeting in supermodular games," *IEEE Transactions on Automatic Control*, vol. 67, no. 12, pp. 6366–6380, 2022.
[10] S. Messina, G. Como, S. Durand, and F. Fagnani, "Optimal intervention in non-binary super-modular games," *IEEE Control Systems Letters*, vol. 7, pp. 2353–2358, 2023.
[11] F. Parise and A. Ozdaglar, "Graphon games: A statistical framework for network games and interventions," *Econometrica*, vol. 91, no. 1, 2023.
[12] M. Granovetter, "Threshold models of collective behavior," *American journal of sociology*, vol. 83, no. 6, pp. 1420–1443, 1978.
[13] W. S. Rossi, G. Como, and F. Fagnani, "Threshold models of cascades in large-scale networks," *IEEE Transactions on Network Science and Engineering*, vol. 6, no. 2, pp. 158–172, 2017.
[14] M. O. Jackson, *Social and Economic Networks*. Princeton, NJ, USA: Princeton University Press, 2008.
[15] C. Ravazzi, G. Como, M. Garetto, E. Leonardi, and A. Tarable, "Asynchronous semianonymous dynamics over large-scale networks," *SIAM Journal on Applied Dynamical Systems*, vol. 22, no. 2, pp. 1300–1343, 2023.
[16] L. Arditti, G. Como, and F. Fagnani, "On the separability of functions and games," *IEEE Transactions on Control of Network Systems*, 2024.
[17] D. M. Topkins, "Equilibrium points in nonzero-sum n-person submodular games," *SIAM Journal on Control and Optimization*, vol. 17, no. 6, pp. 773–787, 1979.
[18] S. Morris, "Contagion," *The Review of Economic Studies*, vol. 67, no. 1, pp. 57–78, 2000.
[19] L. Arditti, G. Como, F. Fagnani, and M. Vanelli, "Robust coordination of linear threshold dynamics on directed weighted networks," *IEEE Transactions on Automatic Control*, 2024.
[20] R. van der Hofstad, *Random Graphs and Complex Networks*. Cambridge University Press, 2016.