Social Nudge with Peer Effects Model

April 20, 2021

Steady State Analysis

This diffusion of social nudges imposes a unique challenge in estimating social nudges' benefit of boosting production in the counterfactual scenario in which *every* user can send and receive nudges. We refer to this as the *global effect* of social nudges. Although it may be tempting to simply compute the global effect as the product of per nudge boosting effect multiplied by the number of nudges sent by users per unit time, the latter term is difficult to quantify due to the diffusion effect. In particular, we notice that both the number of nudges sent per edge from users in the experiment condition and that from users in the control condition underestimate of the number of nudges sent per edge from users when everyone can both send and receive nudges. Blocking nudges for providers in the treatment condition of the experiment naturally hinders the diffusion of nudges. Additionally, this estimation approach ignores the effect of nudges accumulated over time. To tackle these challenges, we propose a novel social network model to capture both the direct effect of social nudges on productivity, and the diffusion of nudges. Applying this model allows us to quantify both the direct and indirect effects of social nudges and estimate the global effect.

We model Platform O as a social network, denoted as $G = (V, E)$, in which $V := \{1, 2, 3, \ldots, |V|\}$ is the set of nodes (i.e., viewers and providers on Platform O) and $E := \{1, 2, 3, \ldots, |E|\}$ is the set of directed edges (i.e., the following relationship on Platform O). We use i, j and e, ℓ to denote nodes and edges, respectively. Let e_o and e_d be the origin and destination, respectively, of edge $e \in E$, so viewer i following provider j is represented as $e_o = i$ and $e_d = j$. We model the dynamics of social nudges and their effects on productivity using a discrete-time stochastic model with infinite time horizon. We use t to index the discrete time period, i.e., $t = 1, 2, \ldots$. In particular, one may consider $t = 1$ as the period when the "social nudge" product function becomes available to all users on the platform. In our empirical context, we assume that each period corresponds to a day, consistent with the business practice of Platform O.

We now model the direct effect of social nudges on production. Let $x_i(t)$ denote the boost of provider i's production in period t due to the social nudges she receives before and during period t. We also use $y_e(t)$ for the number of nudges sent on edge e. Let p_e denote the impact of receiving one social nudge from viewer e_o on the expected production quantity of provider e_d on the day the nudge is received. It is empirically observed that the boosting effect of productivity quickly wears off time. Then, the dynamic of production increment is captured by the dynamic equations

$$x_i(t) = \sum_{1 \le s \le t} \alpha_p^{t-s} \sum_{e \in E: e_d = i} p_e y_e(s), \ \forall i \in V. \tag{1}$$

Here we let $\alpha_p \in (0, 1)$ denote the time discounting factor of the direct production boosting effect.

We next focus on the diffusion of social nudges. Given any $e \in E$, we denote the number of social nudges sent on edge e in period t by $y_e(t)$. We assume that the number of social nudge sent on an edge e in period t are driven by two additive factors. First, a viewer sends nudges to the provider she follows with a base nudging rate not affected by the number of nudges she receives. Second, the diffusion effect dictates that when a provider receives a nudge, she tends to send more nudges out on the social network. Combined, the dynamic of social nudges on G is captured by

$$y_e(t) = \mu_e + \sum_{1 \le s \le t} \alpha_d^{t-s} \sum_{\ell \in E: \ell_d = e_o} d_{\ell e} y_\ell(s) + \epsilon_e^y(t), \ \forall e \in E. \tag{2}$$

Here μ_e represents the base nudging rate and second term embodies the diffusion effect. In particular, $d_{\ell e}$ captures the intensity of social nudge diffusion, i.e., the expected increase of the number of nudges sent on edge e due to one nudge provider e_o receives on edge ℓ directing to e_o (i.e. $\ell_d = e_o$). Note that $d_{\ell e}$ can potentially be heterogeneous across $(\ell, e) \in E^2$. Similar to α_p, $\alpha_d \in (0, 1)$ denotes the time discounting factor of nudge diffusion. We denote the random noise of social nudges sent on edge e in period t as $\epsilon_e^y(t)$, i.i.d. across different periods and edges with a zero mean and a bounded support.

To quantify the global effect of social nudges, we characterize the long-run steady state of the system defined by (1) - (2). Theorem 1 below shows that the expected production and nudge quantities converge to a well-defined and estimable limit. We provide a detailed proof of Theorem 1 in the Appendix. For notation convenience, we define $d_{\ell e} = 0$ if ℓ or e is not in E or $\ell_d \neq e_o$ and matrix $\mathbf{D} := (d_{\ell e} : (\ell, e) \in V^2)$. Matrix $\mathbf{D}$, whose components are all non-negative, therefore captures the first-order diffusion on all edge pairs. We further let $\eta_e := p_e/(1-\alpha_p)$ and $\boldsymbol{\eta} := (\eta_e : e \in V)$. We use $\mathbf{I}$ to denote the identity matrix and $\boldsymbol{e}$ to denote the vector of ones. Let us write the total production increment in period t as $x(t) := \sum_{i \in V} x_i(t)$. We denote ℓ_q-norm of matrices by $||\cdot||_q$ for any $q \in [1, +\infty]$, which is the operator norm defined through $||\mathbf{A}||_q = \sup_{\boldsymbol{z}:||\boldsymbol{z}||_q \le 1} ||\mathbf{A}\boldsymbol{z}||_q$ for any squared matrix $\mathbf{A}$ and $\boldsymbol{z}$ with appropriate dimensions [Horn and Johnson, 2012]. Also, we say that the matrix $\mathbf{D}$ satisfies $\mathcal{C}_q(\delta)$ for some $\delta \in (0,1)$, provided that $||1/(1-\alpha_d)\mathbf{D}||_q \le \delta$.

Theorem 1. *If matrix $\mathbf{D}$ satisfies $\mathcal{C}_q(\delta)$ for some $q \in [1, +\infty]$ and $\delta \in (0, 1)$, it then follows that $\lim_{t\to\infty} \mathbb{E}[x(t)] = x^*$ and $\lim_{t\to\infty} \mathbb{E}[\boldsymbol{y}(t)] = \boldsymbol{y}^*$, where x^* and $\boldsymbol{y}^*$ satisfy $x^* = \boldsymbol{\eta}^\top \boldsymbol{y}^*$ and*

$$\boldsymbol{y}^* = \left(\mathbf{I} - \frac{1}{1-\alpha_d}\mathbf{D}\right)^{-1} \boldsymbol{\mu}. \tag{3}$$

We remark that the condition $||\mathbf{D}||_q < 1 - \alpha_d$ guarantees that $\mathbf{I} - 1/(1-\alpha_d)\mathbf{D}$ is invertible. In this case, (3) can be further expanded and admits a natural interpretation of Bonacich centrality that is reminiscent of some past studies in network economics such as **?**. In this literature, Bonacich centrality is defined for nodes on a network, whereas our centrality measure is for edges. The factors $1/(1-\alpha_d)$ in Eq.(3) and $1/(1-\alpha_p)$ in the definition of $\boldsymbol{\eta}$ materialize the diffusion and productivity boosting effect, respectively, accumulated overtime. To summarize, Theorem 1 characterizes the global effect in the steady state of the production and nudge diffusion process. In the following sections, we discuss in depths how to evaluate the global effect x^* via an approximation scheme and our field experiment.

In our discussion we focus on the production increment x^* in the steady state rather than the total production quantity, but for completeness, we also mention that one can model the expected production of user i in period t as $z_i(t) = \xi_i + x_i(t) + \epsilon_i^z(t)$ for all $i \in V$. Here we denote the provider i's expected production quantity in each period as ξ_i and use $\epsilon_i^z(t)$ to capture a zero mean random noise. In this case, by Theorem 1, if we let the total production quantity in period t as $z(t) := \sum_{i \in V} z_i(t)$, it follows that $\lim_{t\to\infty} \mathbb{E}[z(t)] = \boldsymbol{e}^\top \boldsymbol{\xi} + x^*$.

Approximate Global Effect of Social Nudges

To evaluate the global effect of social nudges on provider productivity, by Theorem 1, we need to invert the $|E|^2$−dimensional matrix $\mathbf{I} - 1/(1-\alpha_d)\mathbf{D}$. For Platform O, the dimension of $\mathbf{D}$, i.e., $|E|^2$, is at the magnitude of 10^{32}. Inverting such a high-dimensional matrix is computationally infeasible. Therefore, we resort to an approximation scheme with a provable performance guarantee to quantify the steady-state (daily) number of social nudges between the viewers and providers, i.e., $\boldsymbol{y}^*$, and the (daily) productivity boost from these nudges,i.e., $\boldsymbol{x}^*$.

Towards this goal, we note that if $\mathbf{D}$ satisfies $\mathcal{C}_q(\delta)$ for some $\delta \in (0,1)$, the inverse of $\mathbf{I} - (1/(1-\alpha_d))\mathbf{D}$ is given by [see, e.g., Corollary 5.6.16 and Corollary 5.6.17 of Horn and Johnson, 2012],

$$\left(\mathbf{I} - \frac{1}{1-\alpha_d}\mathbf{D}\right)^{-1} = \mathbf{I} + \sum_{k=1}^{+\infty} \frac{1}{(1-\alpha_d)^k}\mathbf{D}^k. \tag{4}$$

Motivated by this formula, we define a sequence of (approximate) steady-state social nudge vectors, $\tilde{\boldsymbol{y}}(k)$, and total productivity boosts from nudges, $\tilde{x}(k)$, indexed by $k \in \mathbb{Z}_+$, as follows:

$$\tilde{\boldsymbol{y}}(k) := \left(\mathbf{I} + \sum_{i=1}^{k} \frac{1}{(1-\alpha_d)^i}\mathbf{D}^i\right)\boldsymbol{\mu}, \text{ and } \tilde{x}(k) := \boldsymbol{\eta}^\top \tilde{\boldsymbol{y}}(k).$$

By (4) and Theorem 1, under condition $\mathcal{C}_q(\delta)$, $\tilde{\boldsymbol{y}}(k)$ converges to $\boldsymbol{y}^*$ and $\tilde{x}(k)$ converges to x^* as k approaches infinity. We now characterize the gaps of $\tilde{\boldsymbol{y}}(k)$ and $\tilde{x}(k)$.

Theorem 2. *Assume that $\delta \in (0,1)$ and $\mathbf{D}$ satisfies $\mathcal{C}_q(\delta)$ for some $q \in [1,\infty]$.*

(a) It holds that $\tilde{y}_j(k)$ is increasing in k for any $j \in E$, so is $\tilde{x}(k)$ increasing in k. Therefore, for each $k \in \mathbb{Z}_+$, $\tilde{\boldsymbol{y}}(k)$ is a (component-wise) lower bound of $\boldsymbol{y}^*$, and $\tilde{x}(k)$ is a lower bound of x^*.

(b) For each $k \in \mathbb{Z}_+$, $||\boldsymbol{y}^* - \tilde{\boldsymbol{y}}(k)||_q \le \delta^{k+1}/(1-\delta)||\boldsymbol{\mu}||_q$. Furthermore, $0 \le x^* - \tilde{x}(k) \le \delta^{k+1}/(1-\delta)||\boldsymbol{\mu}||_q||\boldsymbol{\eta}||_r$, where $1/q + 1/r = 1$ with the convention $1/+\infty = 0$.

The key implication of Theorem 2 is that, if the diffusion intensity is not too large (quantified by condition $\mathcal{C}_q(\delta)$), our approximate social nudge vector $\tilde{\boldsymbol{y}}(k)$ and approximate productivity boost $\tilde{x}(k)$ serve as lower bounds, and high-quality approximation of steady-state quantities, with a gap decaying exponentially in k. We also emphasize that, due to the extremely high dimension of the matrix $\mathbf{D}$, evaluating $\tilde{\boldsymbol{y}}(k)$ and $\tilde{x}(k)$ for $k \ge 2$ is also computationally infeasible. For $k = 1$, Theorem 2 shows that the relative gap of our approximation scheme is bounded by $\delta^2||\boldsymbol{\mu}||_1||\boldsymbol{\eta}||_\infty/((1-\delta)x^*)$ (resp. $\delta^2||\boldsymbol{\mu}||_\infty||\boldsymbol{\eta}||_1/((1-\delta)x^*)$) if $\mathbf{D}$ satisfies $\mathcal{C}_1(\delta)$ (resp. $\mathcal{C}_\infty(\delta)$). For Platform O, we leverage the field experiment to estimate the values of δ, $|\boldsymbol{\mu}|_1$, $|\boldsymbol{\mu}|_\infty$, $|\boldsymbol{\eta}|_1$, and $|\boldsymbol{\eta}|_\infty$[1]. We remark that $|\boldsymbol{\mu}|_1$ (resp. $|\boldsymbol{\mu}|_\infty$) is the total (resp. maximum) base nudging rate of all edges on G. Likewise, $|\boldsymbol{\eta}|_1$ (resp. $|\boldsymbol{\eta}|_\infty$) is the total (resp. maximum) productivity boost from nudges for all edges on G.

In fact, at the scale of Platform O, even evaluating $\tilde{\boldsymbol{y}}(1)$ and $\tilde{x}(1)$ involves matrix multiplications in extremely high dimensions, which is again computationally intractable. Therefore, we adopt another layer of approximation by downsampling a subset of providers from V (denoted as $\tilde{V}$). Specifically, we estimate the total productivity boost of all the social nudges received by providers in $\tilde{V}$, denoted as $\hat{\boldsymbol{w}}_0$, and the total productivity boost of the social nudges sent by providers in $\tilde{V}$ *as a result of* the social nudges received by them (i.e., the diffusion of nudges), denoted as $\hat{\boldsymbol{w}}_1$.

[1]To protect sensitive data, we cannot report the true estimates of these values but discuss the estimation method in Appendix xx.

Algorithm 1 APPROXIMATE GLOBAL EFFECT OF SOCIAL NUDGES

Down Sampling: Uniformly randomly sample a subset of nodes $\tilde{V} \subset V$. Find the set of edges that point to a node in $\tilde{V}$, $\tilde{E} := \{e \in E : e_d \in \tilde{V}\}$), and the set of edges that originate from a node in $\tilde{V}$, $\tilde{L} := \{\ell \in E : \ell_o \in \tilde{V}\}$).

Parameter Initialization: For each $e \in \tilde{E}$, estimate μ_e and p_e. For each $\ell \in \tilde{L}$, estimate p_ℓ. For each $e \in \tilde{E}, \ell \in \tilde{L}, e_d = \ell_o$, estimate $d_{e,\ell}$. Estimate α_d and α_p.

Direct Effect of Social Nudges: Estimate

$$\hat{\boldsymbol{w}}_0 := \sum_{i \in \tilde{V}} \sum_{e \in \tilde{E}, e_d = i} \frac{\mu_e p_e}{1 - \alpha_p}.$$

Indirect Effect of Social Nudges: Estimate:

$$\hat{\boldsymbol{w}}_1 := \sum_{i \in \tilde{V}} \sum_{e \in \tilde{E}, \ell \in \tilde{L}, e_d = \ell_o = i} \frac{\mu_e d_{e\ell} p_\ell}{(1 - \alpha_p)(1 - \alpha_d)}$$

Total Productivity Boost on the Entire Population: Scaling the estimates back to V:

$$\hat{\boldsymbol{w}} := \frac{|V|}{|\tilde{V}|}(\hat{\boldsymbol{w}}_0 + \hat{\boldsymbol{w}}_1)$$

Hence, $\hat{\boldsymbol{w}}_0$ captures the direct effect of social nudges and $\hat{\boldsymbol{w}}_1$ captures the indirect effect. Scaling these estimates by a factor of $\frac{|V|}{|\tilde{V}|}$ would therefore yield unbiased estimates of the true direct and indirect global effects. Therefore, we devise $\frac{|V|}{|\tilde{V}|}(\hat{\boldsymbol{w}}_0 + \hat{\boldsymbol{w}}_1)$ as the (unbiased) estimate for $\tilde{\boldsymbol{x}}(1)$. We present a detailed estimation procedure in Algorithm 1.

To obtain the estimate $\hat{\boldsymbol{w}}$, we need to initialize Algorithm 1 with the parameters $\{\mu_e, p_e : e \in \tilde{E}\}$, $\{\gamma_\ell : \ell \in \tilde{L}\}$, $\{d_{e\ell} : e \in \tilde{E}, \ell \in \tilde{L}, e_d = \ell_o\}$, α_p, and α_d.

Appendix 1: Proof of Theorem 2

Let us assume throughout the proof that $\mathbf{D}$ satisfies $C_q(\delta)$ for $q \in [1, \infty]$. The proof follows for any $q \in [1, \infty]$ because we only rely on the sublinearity and submultiplicativity of operator norms. Recall that the system of the social nudge model is defined by the following equations:

$$y_e(t) = \mu_e + \sum_{1 \le s \le t} \alpha_d^{t-s} \sum_{\ell \in E: \ell_d = e_o} d_{\ell e} y_\ell(s) + \epsilon_e^y(t),$$

for all $e \in E$, and for all $i \in V$

$$x_i(t) = \sum_{1 \le s \le t} \alpha_p^{t-s} \sum_{e \in E: e_d = i} p_e y_e(s).$$

Let us denote by $\boldsymbol{y}^*(t) := \mathbb{E}[\boldsymbol{y}(t)]$. Since $\epsilon_e^y(t)$ is the random error with a zero mean and a finite support, it then follows that

$$\boldsymbol{y}^*(t) = \boldsymbol{\mu} + \sum_{1 \le s \le t-1} \alpha_d^{t-s} \mathbf{D} \boldsymbol{y}^*(s) + \mathbf{D} \boldsymbol{y}^*(t),$$

or equivalently

$$\boldsymbol{y}^*(t) = (\mathbf{I} - \mathbf{D})^{-1} \left(\boldsymbol{\mu} + \sum_{1 \le s \le t-1} \alpha_d^{t-s} \mathbf{D} \boldsymbol{y}^*(s) \right). \tag{5}$$

Before we directly focus on $\boldsymbol{y}^*(t)_{t=1}^{+\infty}$, we truncate the number of lags for the ease of argument. That is, we consider the following system with a fixed positive integer M:

$$y_e^M(t) = \mu_e + \sum_{\max\{t-M, 1\} \le s \le t} \alpha_d^{t-s} \sum_{\ell \in E: \ell_d = e_o} d_{\ell e} y_\ell(s) + \epsilon_e^y(t),$$

for all $e \in E$, and for all $i \in V$

$$x_i^M(t) = \sum_{\max\{t-M, 1\} \le s \le t} \alpha_p^{t-s} \sum_{e \in E: e_d = i} p_e y_e(s) q_i.$$

Let us also define $\boldsymbol{y}^{*,M}(t) := \mathbb{E} \boldsymbol{y}^M(t)$. Then, it follows that

$$\boldsymbol{y}^{*,M}(t) = (\mathbf{I} - \mathbf{D})^{-1} \left(\boldsymbol{\mu} + \sum_{\max\{t-M, 1\} \le s \le t-1} \alpha_1^{t-s} \mathbf{D} \boldsymbol{y}^*(s) \right), \tag{6}$$

provided that $\mathbf{I} - \mathbf{D}$ is invertible. Let the spectral radius of a matrix be $\rho(\cdot)$. Then $\rho(\mathbf{D}) \le ||\mathbf{D}||_q < 1 - \alpha_d < 1$, in which the first inequality is due to the fact that the spectral radius of any square matrix is bounded by any operator norm of the matrix. As a result, $\mathbf{I} - \mathbf{D}$ is invertible and $(\mathbf{I} - \mathbf{D})^{-1} = \sum_{i=0}^{\infty} \mathbf{D}^i$ (e.g., see Horn and Johnson [2012]).

Next, we present a series of technical lemmas towards the proof of Theorem 1.

Lemma 3. *The following results regarding $\{\boldsymbol{y}^{*,M}(t)\}_{t=1}^{+\infty}$ holds.*

(*a*) *It follows that $\{\boldsymbol{y}^{*,M}(t)\}_{t=1}^{+\infty}$ is a Cauchy sequence for any M such that $1 \le M < \infty$.*

(*b*) *There exists a positive constant θ such that $||\boldsymbol{y}^{*,M}(t)||_q \le \theta$ and $||\boldsymbol{y}^*(t)||_q \le \theta$ for any $t \in \mathbb{Z}_+$ and $M \in \mathbb{Z}_+$.*

Note that a Cauchy sequence is always bounded, but the second term of Lemma 3 implies these sequences are uniformly bounded over any M, including when $M = \infty$. This fact is useful in the proof of Lemma 4, but slightly strong than necessary. Now let us turn to the sequence $\{\boldsymbol{y}^*(t)\}_{t=1}^{+\infty}$. We state the following lemma.

Lemma 4. *For any $\epsilon > 0$, there exists $\bar{M}$ such that if $M \geq \bar{M}$, $||\boldsymbol{y}^*(t) - \boldsymbol{y}^{*,M}(t)||_q \leq \epsilon$ uniformly for t.*

Lemmas 3 and 4 together imply that $\{\boldsymbol{y}^*(t)\}_{t=1}^{+\infty}$ is also a Cauchy sequence, and therefore the limit of the sequence $\{\boldsymbol{y}^*(t)\}_{t=1}^{+\infty}$ exists. In particular, given any ϵ, there exists M and T such that $||\boldsymbol{y}^*(t) - \boldsymbol{y}^{*,M}(t)||_q \leq \epsilon/3$ for any t, and $||\boldsymbol{y}^{*,M}(t_1) - \boldsymbol{y}^{*,M}(t_2)||_q \leq \epsilon/3$ for any $t_1, t_2 \geq T$. Consequently, it holds that

$$||\boldsymbol{y}^*(t_1) - \boldsymbol{y}^*(t_2)||_q \leq ||\boldsymbol{y}^*(t_1) - \boldsymbol{y}^{*,M}(t_1)||_q + ||\boldsymbol{y}^{*,M}(t_1) - \boldsymbol{y}^{*,M}(t_2)||_q + ||\boldsymbol{y}^{*,M}(t_2) - \boldsymbol{y}^*(t_2)||_q \leq \epsilon,$$

and $\{\boldsymbol{y}^*(t)\}_{t=1}^{+\infty}$ is Cauchy, whose limit we write as $\boldsymbol{y}^*$. Let us note that by Eq.(5)

$$\boldsymbol{y}^*(t+1) = (\mathbf{I} - \mathbf{D})^{-1}\left(\boldsymbol{\mu} + \sum_{1 \leq s \leq t} \alpha_d^{t+1-s}\mathbf{D}\boldsymbol{y}^*(s)\right) \tag{7}$$

and

$$\alpha_1 \boldsymbol{y}^*(t) = (\mathbf{I} - \mathbf{D})^{-1}\left(\alpha_d\boldsymbol{\mu} + \sum_{1 \leq s \leq t-1} \alpha_d^{t-s+1}\mathbf{D}\boldsymbol{y}^*(s)\right). \tag{8}$$

Taking the differnce between Eq.(7) from Eq.(8) leads to

$$\boldsymbol{y}^*(t+1) - \alpha_d \boldsymbol{y}^*(t) = (\mathbf{I} - \mathbf{D})^{-1}\left((1-\alpha_d)\boldsymbol{\mu} + \alpha_d \mathbf{D}\boldsymbol{y}^*(t)\right).$$

Letting $t \to +\infty$ on both sides leads to $\boldsymbol{y}^* - \alpha_d \boldsymbol{y}^* = (\mathbf{I} - \mathbf{D})^{-1}\left((1-\alpha_d)\boldsymbol{\mu} + \alpha_d \mathbf{D}\boldsymbol{y}^*\right)$. Reorganizing the terms, we have $\boldsymbol{y}^* = \boldsymbol{\mu} + 1/(1-\alpha_d)\mathbf{D}\boldsymbol{y}^*$. Therefore, $\boldsymbol{y}^* = (\mathbf{I} - 1/(1-\alpha_d)\mathbf{D})^{-1}\boldsymbol{\mu}$, because the spectral radius of $1/(1-\alpha_d)\mathbf{D}$, i.e., $\rho(1/(1-\alpha_d)\mathbf{D})$, satisfies

$$\rho\left(\frac{1}{1-\alpha_d}\mathbf{D}\right) \leq \left|\left|\frac{1}{1-\alpha_d}\mathbf{D}\right|\right|_q < 1$$

because of our assumption. The proofs of $\lim_{t\to\infty} \mathbb{E}\boldsymbol{x}(t) = \boldsymbol{x}^*$ and

$$x_i^* = \frac{1}{1-\alpha_p} \sum_{e \in E: e_d = i} p_e y_e^*,$$

for all $i \in V$ are similar to our arguments so far. Therefore, we omit this part and conclude the proof. □

In the next, we prove several technical lemmas. The following fact is useful.

Lemma 5. *It holds that $(\alpha_d/(1-\alpha_d))||(\mathbf{I} - \mathbf{D})^{-1}\mathbf{D}||_q < 1$.*

Proof of Lemma 3. Towards the proof of item (a) in Lemma 3, let us fix M. Consider two real number sequences $\{\boldsymbol{y}^1(t)\}_{t=1}^M$ and $\{\boldsymbol{y}^2(t)\}_{t=1}^M$, and let

$$\bar{\boldsymbol{z}}^i = (\mathbf{I} - \mathbf{D})^{-1}\left[\boldsymbol{\mu} + \sum_{1 \leq s \leq M} \alpha_d^{M-s+1}\mathbf{D}\boldsymbol{y}^i(s)\right],$$

for $i = 1, 2$. It then follows that

$$||\bar{\boldsymbol{z}}^1 - \bar{\boldsymbol{z}}^2||_q = \left|\left|(\mathbf{I} - \mathbf{D})^{-1} \sum_{1 \leq s \leq M} \alpha_d^{M-s+1} \mathbf{D} \Big(\boldsymbol{y}^1(s) - \boldsymbol{y}^2(s) \Big) \right|\right|_q$$

$$\leq \big|\big|(\mathbf{I} - \mathbf{D})^{-1}\mathbf{D}\big|\big|_q \left|\left| \sum_{1 \leq s \leq M} \alpha_d^{M-s+1} \Big(\boldsymbol{y}^1(s) - \boldsymbol{y}^2(s) \Big) \right|\right|_q = \big|\big|(\mathbf{I} - \mathbf{D})^{-1}\mathbf{D}\big|\big|_q \cdot \sum_{1 \leq s \leq M} \alpha_d^{M-s+1} \cdot \left|\left| \Big(\hat{\boldsymbol{y}}^1 - \hat{\boldsymbol{y}}^2 \Big) \right|\right|_q$$

$$\leq \big|\big|(\mathbf{I} - \mathbf{D})^{-1}\mathbf{D}\big|\big|_q \left|\left| \cdot \sum_{1 \leq s \leq M} \alpha_d^{M-s+1} \cdot \mathbf{D} \right|\right|_q \max_{1 \leq s \leq M} \left|\left| \Big(\boldsymbol{y}^1(s) - \boldsymbol{y}^2(s) \Big) \right|\right|_q \tag{9}$$

in which the first inequality follows due to the sub-multiplicity of ℓ_2-norm and the second equality follows from definition

$$\hat{\boldsymbol{y}}^i := \Big(\sum_{1 \leq s \leq M} \alpha_d^{M-s+1} \boldsymbol{y}^i(s) \Big) \Big/ \Big(\sum_{1 \leq s \leq M} \alpha_d^{M-s+1} \Big),$$

and the sub-mulplicity of ℓ_2-norm, and the last inequality follows from the convexity of ℓ_2-norm. By Lemma 5, $(\alpha_d/(1-\alpha_d))||(\mathbf{I} - \mathbf{D})^{-1}\mathbf{D}||_q = c$ for some constant $c < 1$. Therefore, since $\sum_{1 \leq s \leq M} \alpha_d^{M-s+1} \leq \alpha_d/(1-\alpha_d)$, we observe

$$||\bar{\boldsymbol{z}}^1 - \bar{\boldsymbol{z}}^2||_q \leq c \cdot \max_{1 \leq s \leq M} \left|\left| \Big(\boldsymbol{y}^1(s) - \boldsymbol{y}^2(s) \Big) \right|\right|_q. \tag{10}$$

As a result of (10), if we define $w(t) = ||\boldsymbol{y}^{*,M}(t+1) - \boldsymbol{y}^{*,M}(t)||_q$, for $k \geq 1$ it holds

$$w(kM+1) \leq c \cdot \max_{1 \leq s \leq M} \{w((k-1)M+s)\}.$$

Similarly, $w(kM+2) \leq c \cdot \max \Big\{ \max_{2 \leq s \leq M}\{w((k-1)M+s)\}, w(kM+1) \Big\} \leq c \cdot \max_{1 \leq s \leq M}\{w((k-1)M+s)\}$, in which the second inequality follows from $w(kM+1) \leq c \cdot \max_{1 \leq s \leq M}\{w((k-1)M+s)\}$. Induction shows that $w(kM+s) \leq c \cdot \max_{1 \leq t \leq M}\{w((k-1)M+t)\}$ for all s satisfy $1 \leq s \leq M$. Let us define $\tau = \max_{1 \leq s \leq M}\{w(s)\}$. By induction one can verify that

$$w(kM+s) \leq c^k \tau,$$

for all $k \geq 0$ and $1 \leq s \leq M$. This implies that for some k, k', s and s' such that $Mk'+s' \geq Mk+s$, we have

$$||\boldsymbol{y}^{*,M}(Mk'+s') - \boldsymbol{y}^{*,M}(Mk+s)||_q \leq \sum_{i=Mk+s}^{Mk'+s'-1} ||\boldsymbol{y}^{*,M}(i+1) - \boldsymbol{y}^{*,M}(i)||_q \leq \sum_{i=2Mk+s}^{+\infty} ||\boldsymbol{y}^{*,M}(i+1) - \boldsymbol{y}^{*,M}(i)||_q$$

$$= \sum_{i=Mk+s}^{+\infty} w(i) \leq \sum_{j=0}^{\infty} Mc^{k+j}\tau = Mc^k \cdot \frac{\tau}{1-c},$$

where the first inequality follows from triangular inequality. This implies that $\{\boldsymbol{y}^{*,M}(t)\}_{t=1}^{+\infty}$ is a Cauchy sequence. This finishes the proof of item (a) in Lemma 3.

Let us recall from Lemma 5 that $\alpha_d/(1-\alpha_d)\cdot||(\mathbf{I}-\mathbf{D})^{-1}\mathbf{D}||_q = c$ for some constant $c<1$. Let us choose a positive constant θ satisfying $||(\mathbf{I}-\mathbf{D})^{-1}\boldsymbol{\mu}||_q + c\theta \le \theta$ and $||\boldsymbol{y}^{*,M}(1)||_q \le \theta$ for all $M \le +\infty$ (note that we understand $\boldsymbol{y}^{*,\infty}(t)$ as $\boldsymbol{y}^{*,\infty}(t) = \boldsymbol{y}^*(t)$). This choice of θ is possible since $||\boldsymbol{y}^{*,M}(1)||_q = ||\boldsymbol{y}^{*,M'}(1)||_q$ for any M, M'. Furthermore, for any t and M such that $1 \le t, M \le +\infty$, we argue that $||\boldsymbol{y}^{*,M}(t)||_q \le \theta$ by induction. Fix M and let us assume that this holds that any $s<t$. It then holds by Eq.(6) that,

$$\begin{aligned}
||\boldsymbol{y}^{*,M}(t)||_q \le & ||(\mathbf{I}-\mathbf{D})^{-1}\boldsymbol{\mu}||_q + \sum_{\max\{t-M,1\}\le s\le t-1} \alpha_d^{t-s} \left|\left|(\mathbf{I}-\mathbf{D})^{-1}\mathbf{D}\boldsymbol{y}^*(s)\right|\right|_q \\
\le & ||(\mathbf{I}-\mathbf{D})^{-1}\boldsymbol{\mu}||_q + \sum_{\max\{t-M,1\}\le s\le t-1} \alpha_d^{t-s} \left|\left|(\mathbf{I}-\mathbf{D})^{-1}\mathbf{D}\right|\right|_q ||\boldsymbol{y}^*(s)||_q \\
\le & ||(\mathbf{I}-\mathbf{D})^{-1}\boldsymbol{\mu}||_q + \frac{\alpha_d}{1-\alpha_d}\left|\left|(\mathbf{I}-\mathbf{D})^{-1}\mathbf{D}\right|\right|_q \theta \le ||(\mathbf{I}-\mathbf{D})^{-1}\boldsymbol{\mu}||_q + c\theta \le \theta,
\end{aligned}$$

in which the first inequality follows form the sublinearity of ℓ_2-norm, the second inequality follows from the submulplicity of ℓ_2-norm, the third inequality follows from the geometric sum, the fourth inequality follows from the definition of c, and the last inequality follows from the definition of θ. This concludes the proof of the second item. □

Proof of Lemma 4. Clearly, $||\boldsymbol{y}^*(t) - \boldsymbol{y}^{*,M}(t)||_q = 0$ for $t \le M+1$ by Eq.(6). Assume that $t > M+1$ and $||\boldsymbol{y}^*(s) - \boldsymbol{y}^{*,M}(s)||_q \le \epsilon$ holds for all $s<t$. By the subadditivity of ℓ_2-norm, we have

$$\begin{aligned}
& ||\boldsymbol{y}^*(t) - \boldsymbol{y}^{*,M}(t)||_q \\
\le & \sum_{1\le s<t-M} \alpha_d^{t-s}\left|\left|(\mathbf{I}-\mathbf{D})^{-1}\mathbf{D}\boldsymbol{y}^*(s)\right|\right|_q + \left|\left|(\mathbf{I}-\mathbf{D})^{-1}\cdot \sum_{\max\{t-M,1\}\le s\le t-1}\alpha_d^{t-s}\mathbf{D}\left(\boldsymbol{y}^{*,M}(s)-\boldsymbol{y}^*(s)\right)\right|\right|_q
\end{aligned}$$

The first term goes to zero as $M\to\infty$, because $||\boldsymbol{y}^*(t)||_q$ is bounded by Lemma 3. More specifically, it holds that $\sum_{1\le s<t-M}\alpha_d^{t-s} \le \sum_{s=M}^{\infty}\alpha_d^s = \alpha_d^M/(1-\alpha_d) \to 0$. Therefore, the first term vanishes as M is large.

With an argument similar to Lemma 3, the second term can be bounded by

$$\left|\left|\sum_{\max\{t-M,1\}\le s\le t-1}\alpha_d^{t-s}(\mathbf{I}-\mathbf{D})^{-1}\mathbf{D}\right|\right|_q \max_{t-M\le s\le t-1}\{||\boldsymbol{y}^{*,M}(s)-\boldsymbol{y}^*(s)||_q\} \le c \max_{t-M\le s\le t-1}\{||\boldsymbol{y}^{*,M}(s)-\boldsymbol{y}^*(s)||_q\},$$

which by the induction hypothesis is bounded by $c\epsilon$. Therefore, as long we pick M large enough so that the first term is bounded by $(1-c)\epsilon$, we have $||\boldsymbol{y}^*(t) - \boldsymbol{y}^{*,M}(t)||_q \le \epsilon$. □

Proof of Lemma 5. As argued in the proof of Theorem 1, $||\mathbf{D}|| < 1-\alpha_d$ implies that $\mathbf{I}-\mathbf{D}$ is invertible and $(\mathbf{I}-\mathbf{D})^{-1} = \sum_{i=0}^{\infty}\mathbf{D}^i$ where the convergence is understood element-wise. Therefore,

$$||(\mathbf{I}-\mathbf{D})^{-1}\mathbf{D}||_q = ||\sum_{i=0}^{\infty}\mathbf{D}^i\cdot\mathbf{D}||_q = ||\sum_{i=1}^{\infty}\mathbf{D}^i||_q \le \sum_{i=1}^{\infty}||\mathbf{D}^i||_q \le \sum_{i=1}^{\infty}||\mathbf{D}||_q^i < \sum_{i=1}^{\infty}\alpha_d^i = \frac{\alpha_d}{1-\alpha_d}.$$

Therefore, the conclusion follows. □

Appendix 2: Proof of Theorem 2

To simplify the presentation, we define $\mathbf{\Sigma} := (1/(1-\alpha_d))\mathbf{D}$. We assume throughout the proof condition $\mathbf{D}$ satisfies $\mathcal{C}_q(\delta)$ for some $q \in [1, +\infty]$ and $\delta \in (0,1)$. By (4), we have

$$(\mathbf{I} - \mathbf{\Sigma})^{-1} = \mathbf{I} + \sum_{i=1}^{+\infty} \mathbf{\Sigma}^i.$$

Before giving the proof of Theorem 2, we first introduce a sequence of matrices

$$\{\boldsymbol{A}(k) := \mathbf{\Sigma}^k, k = 1, 2, ...\}.$$

The following lemma is a stepping stone for the proof of Theorem 2.

Lemma 6. *It holds that* $||\mathbf{A}(k)\boldsymbol{\mu}||_q < \delta^k ||\boldsymbol{\mu}||_q$.

Proof. By the sub-multiplicative property of matrix norm,

$$||\mathbf{A}(k)\boldsymbol{\mu}||_q = ||\mathbf{\Sigma}^k \boldsymbol{\mu}||_q \leq (||\mathbf{\Sigma}||_q)^k ||\mu||_q < \delta^k ||\mu||_q.$$

This completes the proof of Lemma 6. □

We are now ready to prove Theorem 2.

Proof of Theorem 2.

For Part(a), by (4), we have $\boldsymbol{y}_e^* = ((\mathbf{I} + \sum_{i=1}^{+\infty} \mathbf{\Sigma}^i)\boldsymbol{\mu})_e$ for each $e \in E$, and $x^* = \boldsymbol{\eta}^T(\mathbf{I} + \sum_{i=1}^{+\infty} \mathbf{\Sigma}^i)\boldsymbol{\mu}$. Since $\mathbf{\Sigma}$ is non-negative, $\boldsymbol{y}_e^* - \tilde{\boldsymbol{y}}_e(k) = \sum_{i=k+1}^{+\infty} (\mathbf{A}(i)\boldsymbol{\mu})_e$ is decreasing in k for $e \in E$. So is $x^* - \tilde{x}(k) = \sum_{i=k+1}^{+\infty} \boldsymbol{\eta}^T \mathbf{A}(i)\boldsymbol{\mu}$ decreasing in k as well. This proves Part (a).

For Part (b), by Lemma 6, we have

$$||\boldsymbol{y}^* - \tilde{\boldsymbol{y}}(k)||_q \leq \sum_{i=k+1}^{+\infty} ||\mathbf{A}(i)\boldsymbol{\mu}||_q < \sum_{i=k+1}^{+\infty} \delta^i ||\boldsymbol{\mu}||_q = \frac{\delta^{k+1}}{1-\delta} ||\boldsymbol{\mu}||_q,$$

where the first inequality follows from subadditivity of matrix norms, and the second inequality follows from Lemma 6. Also,

$$|x^* - \tilde{x}(k)| = |\boldsymbol{\eta}^\top (\boldsymbol{y}^* - \tilde{\boldsymbol{y}}(k))| \leq ||\boldsymbol{\eta}||_r ||\boldsymbol{y}^* - \tilde{\boldsymbol{y}}(k)||_q \leq ||\boldsymbol{\eta}||_r \cdot \frac{\delta^{k+1}}{1-\delta} ||\boldsymbol{\mu}||_q,$$

where the first inequality follows from Hölder's inequality. This concludes the proof of Theorem 2(b). □

References

Roger A Horn and Charles R Johnson. *Matrix analysis*. Cambridge university press, 2012.