Optimization of Cascading Processes in Multiscale Networks with Stochastic Interactions

Juan S. Borrero, Oleg A. Prokopyev
Department of Industrial Engineering, University of Pittsburgh, Pittsburgh, PA 15261, {jsb81,droleg}@pitt.edu

Pavlo Krokhmal
Department of Systems & Industrial Engineering, University of Arizona, Tucson, AZ, 85721, krokhmal@email.arizona.edu

We consider the problem of optimal propagation of cascading, or "domino"-type processes in networks under the presences of uncertainties. To this end, we propose a general model that represents a cascading process in a network as a Markov process on a multiscale graph. Assuming that the parameters governing this Markov process can be modified at certain costs, we investigate the question of optimal resource allocation that would minimize the time by which the cascading process reaches all the nodes in the network. Under some simple assumptions, we derive analytical solutions for an optimal resource allocation that optimizes the spread of the cascade. These expressions explicitly elucidate the importance of network interactions in the context of the rate at which the cascade spreads. We show that an optimal resource allocation can be computed in a strongly polynomial time in terms of the size of the network. Importantly, the obtained solution is expressed via a minimum spanning arborescence on an auxiliary graph and provides a hierarchy that classifies the clusters of the network in terms of their importance with respect to the cascade propagation.

Key words: Cascade propagation, influence propagation, influence maximization, Markov chains, multiscale graphs, minimum spanning arborescence

1. Introduction

Large-scale systems, such as engineered and social networks, often exhibit *cascading* behavior during which the propagation of information, influence, failures, etc., has a global effect on the system albeit starting from few locations (Watts [49]). The study of such processes has been receiving an increasing amount of attention in the recent literature, e.g., in the context of failure/recovery of interconnected engineered systems, see e.g., Amin [3], Duenas-Osorio and Vemuru [18], viral marketing, as in Domingos and Richardson [16], Richardson and Domingos [41], and more generally, in the way ideas and innovations propagate through a social network, see Kempe et al. [31]. In particular, one of the main challenges is associated with understanding of the specific network features that trigger and sustain such effects in these complex settings (Little [35], Watts [49]).

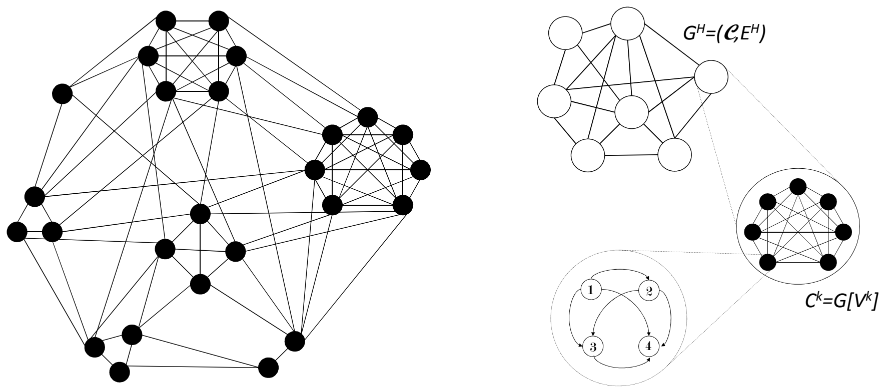
Among the key questions pertaining to cascade processes in networks is one of the control and optimization of such processes subject to resource allocation constraints, such as physical or budgetary constraints. From this perspective, the influence maximization problem, or target-seed selection problem, has received considerable attention from the research community, see, among others, Ackerman et al. [1], Ben-Zwi et al. [8], Chen et al. [15], Kempe et al. [30]. In this problem, inspired mainly by marketing applications in social networks, one has to choose a limited number of nodes to start spreading an idea or a rumor with the objective to maximize the number of influenced nodes through a cascading process. Similarly, Blume et al. [10] study network resiliency optimization models that are concerned with the identification of network topologies that minimize the number of failed nodes after the cascade propagation.

In this paper we propose a class of models that provide both quantitative and qualitative insights into the cascade propagation phenomenon in large complex networks under uncertainties. The primary focus of this work is on the development of models and solution procedures that are amenable to analytical treatment, allowing one to elucidate the structural and qualitative behavior of cascade propagation processes.

The proposed approach to modeling of the large-scale, cascading changes in complex networks in the presence of uncertainties is based on representing the state of the cascade in the network as a stochastic process on a multiscale graph. At the higher scale (see Fig. 1), a large complex network is modeled by a directed weighted graph whose vertices represent dense interconnected clusters of the network. The arcs and weights of the high-scale graph measure the influence that clusters have on each other and depend on the structure of the original network.

The medium-scale networks are the dense clusters and it is assumed that the network nodes at the this scale are the basic units that provide network's functionality, and therefore at any time moment they can be in one of several states of the cascade, e.g., "functioning normally," "failed," "in recovery," etc. The set of states of a medium-scale node is formalized as a lower-scale directed graph, where vertices represent the possible states, and directed edges denote the possible transitions between states (Fig. 1(b)). As the nodes in medium-scale networks change their states not completely deterministically but under influence of various stochastic factors, it is assumed that the state of such a node can be described as a stochastic process on the lower-scale graph. In particular, in this endeavor we adopt the Markovian framework for modeling of the nodes' temporal behavior under uncertainty.

Further, within the presented approach we assume that the Markov processes governing different nodes in the medium-scale graph are interdependent, such that a change of state of one node may influence (e.g., change the parametrization of) the stochastic processes at the other nodes in the network. In this regard, the multiscale decomposition allows us to model "Close" and "Distant" influence relationships. In "Close" (cluster) relationships, the state of a node of a medium-scale graph has a direct effect on the state of the nodes in the same cluster. This phenomenon is observed in social networks, where medium-level graphs can be thought of as cohesive groups of people (e.g., close friends, cliques) whose members share strong ties. In such close groups, the preferences and



(a) Original network, $G = (V, E)$ (b) Multiscale decomposition

Figure 1 The network at the top of (b) is a high-level graph, in which each node corresponds to a cluster (clique) of the network in (a). A node of the medium-level graph in (b) is further described by a lower-level directed graph that demonstrates possible state transitions of each node, as detailed at the bottom of (b). For more detailed discussion on the notation see Section 4.1.

opinions of each member have a direct influence on the preferences of the other group members, see, e.g., Wasserman and Faust [48], Hill et al. [26]. In contrast, the "Distant" (non-cluster) relationships are those in which the state of a node of a medium scale graph has an indirect effect on the state of the cascade of neighboring nodes in other clusters. Particularly, we model this indirect non-cluster influence through an aggregated measure that depends on the size of the clusters, their overall state with respect to the cascade, and the weights of the higher-scale graph. Such "Distant" influence relationship allow us to model the role that groups of nodes have on cascade/influence propagation (Watts and Dodds [50]).

In our model, the parametrization of the stochastic processes governing the cascade at any node reflects three *modes of influence* by which the cascade spreads throughout the network. Namely, for each cluster of the network, these modes are (i) *intrinsic capability*, which stands for a node's ability to progress on its own through the states/stages of the cascade; (ii) *network* or *community influence*, which represents the effect that other nodes in the network have on a node's state during the cascade; (iii) *external influence*, or *external help*, which denotes the influence that each node of the cluster receives from a shared resource outside the network.

The definitions of these modes are motivated by applications. As an example, consider the context of propagating a new product or idea throughout a network of consumers (see, e.g., Domingos and Richardson [16]). Intrinsic capability corresponds to a customer adopting the product by himself/herself, e.g., by seeing it in stores and becoming convinced of its superior qualities, or being influenced by public advertisement campaign, or more generally, by a large 'broadcast' (Goel et al. [23]). Network influence in this setting means influence (via viral marketing or "word-of-mouth" mechanisms) from the consumers who already adopted the product, while the external influence corresponds to a resource that is shared among customers. For example, it may correspond to in-store or door-to-door product demonstrations, discount coupons distributed among customers, etc.

For an alternative interpretation, suppose that the network represents an interconnected system that has been subject to a *large-scale failure*. Initially all nodes are failed, but their recovery is possible and thus the system functionality can be restored (Majdandzic et al. [38]). In this setting, intrinsic capability measures the ability of a node to recover by its own means, network influence corresponds to the assistance that failed nodes receive from nodes that have completely recovered, and external influence represents support that a node receives from a external agent (e.g., a "repairer") exogenous to the network.

Given the parametrizations of the Markov processes that govern individual medium-scale nodes, a natural question is how to select the values of parameters that improve/minimize the time until the cascade propagates across all the nodes in the network. To this end, we assume that the value of a given parameter can be modified at a certain cost, which leads to the problem of optimal resource allocation that minimizes the time until are nodes of the network are subject to the cascade.

Within this framework, we describe the evolution of the cascade in the network by using a continuous time Markov chain (CTMC). The states in this CTMC correspond to the number of nodes in each cluster that are in each state, and the transition rates of the chain are given in terms of the parameters of the stochastic processes that govern the evolution of the cascade at each node. We show how the minimization of the cascade propagation time can be posed as the minimization of the second largest eigenvalue of the generator matrix of the CTMC, which in turn can be framed as linear programming problem (LP).

Under natural assumptions on the relationship between the model's costs and coefficients, we derive closed-form analytical solutions that provide insights into the cascade propagation. First, we characterize how the decision-maker should determine the mode(s) of influence and the effort(s) that has to be exerted by each mode on each cluster of the network. Second, we show that if it is optimal for the decision-maker to invest into network influence, then the cascade rate grows exponentially with the size of the network, thus, demonstrating the domino effect caused by network interactions. If, in contrast, an optimal solution is to invest either into external influence or intrinsic capabilities exclusively (e.g., the links within the network are sufficiently 'weak'), then the cascade spread rate does not depend on the size of the network. Third, we show that there exist certain states of the Markov chain that act as bottlenecks for the spread of influence; which suggest that any improvement of the cascade rate must decrease the time the CTMC spends in those states.

More importantly, we demonstrate a relationship between the cascade propagation optimization problem and a minimum spanning arborescence (MSA) problem on an auxiliary graph. The solution of this MSA problem yields a tree-like hierarchy of the network clusters that classifies them in terms of their importance with respect to the cascade propagation. This classification identifies clusters which can be used as seeds of the cascade process in the underlying network. In particular, our model shows that large clusters should be such seeds in settings where the parameters defining the influence spread are homogenous across the network. Moreover, the values of the analytical solution of the LP depend on the value of the MSA, and as finding a MSA can be done in polynomial time in terms of the size of the network, then an optimal solution of the cascade propagation optimization problem can be found in polynomial time.

2. Related Literature

One of the earlier works on influence propagation optimization is presented by Richardson and Domingos [41], who model cascading effects by using a Markov random field. In this model, the probability that a customer buys a product depends on the decisions of its neighbors and on the attributes of the product being marketed. The decision problem is whether or not to market to a customer, and the resulting optimization problem is solved heuristically.

Kempe et al. [30] introduce the Influence Maximization (IM) problem, in which a fixed number of seed nodes have to be chosen to become activated (i.e., influenced) to initiate propagation. Two probabilistic models for influence are proposed, namely, the linear threshold (LT) model and the independence cascade (IC) model. In the LT model a node becomes active as soon as a linear function that depends on the number of its active neighbors surpasses a threshold. In the IC model a node that has just become active has a single chance to activate each of its neighbors with a certain probability. In both settings the objective is to maximize the number of nodes that are activated after the cascade ends (i.e., the first period when there are no more newly activated nodes). The authors frame the problem as an integer program and solve it approximately by exploiting the submodularity of the objective function.

Within the framework of the IM model, most of the research efforts have been focused on either refining solution methods or considering its modeling variations. Complete overviews on the IM model and its variations can be found in Kempe et al. [31] and Chen et al. [13]; also, an analysis of the citations of this model, and related papers in cascade propagation, is given in Samadi [44]. Below we discuss the variations more relevant to our work.

Chen et al. [14] introduce time considerations into the IM model. They focus on the concept of meeting-events, where the number of periods a newly activated node takes to activate its neighbor is geometrically distributed. Liu et al. [36] generalize this model to account for other types of delay distributions beyond geometric. Goyal et al. [25] consider an alternative optimization problem where they study the problem of minimizing the number of time periods necessary to activate a given number of nodes if a fixed amount of nodes can be activated initially. Gomez Rodriguez et al. [24] propose another variation of the IM model, where an active node triggers the activation of a neighboring node after an exponentially distributed time whose parameter depends on each pair of nodes.

Motivated by the challenges of scaling the IM model for large-scale networks, Eftekhar et al. [20] consider the seed selection optimization problem over clusters of nodes rather than over individual nodes. The model follows the dynamics of the IC model, but instead of focusing on individual activations, a measure related to the expected fraction of activated nodes within each cluster is being tracked. A related approach based on first detecting dense communities within the network, and then finding the most influential nodes over these communities, is presented in Wang et al. [47].

The focus of the literature discussed above is on the numerical and algorithmic issues arising in the optimization problems; in contrast, our model is centered around establishing structural and analytical results. Furthermore, different from most of these studies (with the exception of Eftekhar et al. [20], Wang et al. [47]), we consider a multiscale model for network interactions, in which the activation of a node might explicitly depend on other nodes beyond its neighbors, i.e., there are "Close" (direct) and "Distant" (indirect) influence relationships. We note that our model can be restricted to only represent direct influence relationships (as in the IM model) by considering that each node of the network is a cluster on its own, see discussion in Section 4.1.

Our approach is also different from an optimization perspective, since rather than optimizing over the space of nodes to determine which one of them to activate initially, we optimize over the space of feasible values of a parameter vector; this method yields a more general framework for studying cascade effects. Hence, our model can be more useful for large-scale networks where analyzing generic influence attributes can be more meaningful than determining specific seeds for initial activation. Finally, optimal solutions of our cascade propagation optimization problem can be computed in polynomial time, while the IM models typically require application of complex custom approximation and simulation methods due to *NP*-hardness of the corresponding optimization formulations.

One of the main features of the presented approach is that the distribution of states of each node depends on the current states of other nodes of the network. A similar type of systems is often studied by using the concept of stochastic cellular automata, e.g., Bhargava et al. [9], Plateau and Atif [40]. In particular, Asavathiratham et al. [5] study the propagation of influence through a network by using a generic modeling framework of interactive discrete-time Markov chains referred to as *the influence model*. The focus of their work is on computing the distributions (transient or steady state) and moments of different subsets of interacting chains. A more recent and general treatment of the influence model can also be found in Richoux and Verghese [42], while some related applications are discussed by Dong and Pentland [17], Ma et al. [37], Peng et al. [39]. Interacting Markov chains are also used to study the effect of cueing in military search missions, see e.g., Jeffcoat et al. [27]. We note that with the exception of Roy et al. [43], who consider a particular multi-stage repair resource allocation problem within the influence model framework, to the best of our knowledge no optimization approaches have been considered for systems consisting of interacting Markov chains.

Finally, there is a significant body of related work in the context of failure cascades, see, for example, Brummitt et al. [11], Kim et al. [32], Wu et al. [52], and the references therein. These works, however, do not involve optimization, and typically focus on the performance evaluation of different network classes under particular cascade models. A notable exception is the work by Blume et al. [10], who seek an optimal network configuration so as to alter the cascade propagation.

3. A Cascading Process in Complete Networks

In this section we develop a model for propagation of cascading processes in complete networks (i.e., networks with all possible edges), which serves as the main building block for the analysis of cascades on general networks discussed in Section 4. In this particular case, the high-level graph consists of only one node, and thus the focus of this section is to describe in detail the stochastic dynamics of a medium-level graph. We begin with a motivating example when the lower-level graphs have only two nodes (states). The general model, the corresponding Markov chain framework, and the optimization problem are discussed in Section 3.2. The main properties of the model and its solution, and the analytical insights are discussed in Section 3.3. In Section 3.4, we present an asymptotic analysis of the model's behavior when the number n of nodes in the network is infinitely large.

3.1. An Example of the Model with Two States

Consider a complete medium-scale graph, where each node at any given time may be in one of only two states, which we refer to as *inactive* and *active*, see Figure 2. This (abstract) nomenclature serves the purpose of generality: we may equivalently think about the nodes representing consumers each of whom is either a *customer* or *not a customer* of a product/service, or system units which can be in a *failed* or *recovered* condition, or members of a population who are either *infected* with a virus or *healthy*, and so on. Assume that all nodes are initially in the inactive state and may become active at some point in future (e.g., the consumers who are initially unaware of the new product may in time become its adopters, failed system units may recover from failure, etc). Namely, an inactive node can become active on its own (i.e., via the *intrinsic capability*); due to influence from

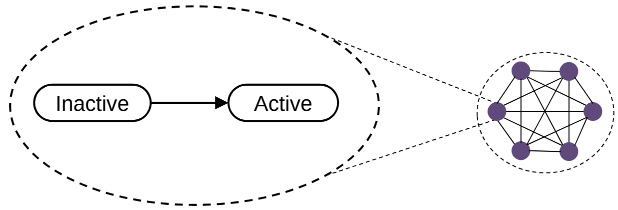


Figure 2 Lower-level graph of the model.

active nodes in the network (i.e., via the *network influence*); and/or by receiving influence from an entity exogenous to the network (i.e., via the *external influence*).

We assume that the time it takes an inactive node to activate on its own is exponentially distributed with the rate θ, and that the time it takes an active node (respectively, the external entity) to activate an inactive node is exponentially distributed with the rate η (respectively, μ). Motivated by the fact that the network is a clique, we assume that as soon as an active node successfully influences a node, it immediately starts influencing any other inactive node. The same holds for the external entity. Note that this implies that a group of active nodes and the external entity might try to influence the same node simultaneously.

Let $Y(\boldsymbol{\zeta}, t)$ be the number of inactive nodes in the network at time t, where $\boldsymbol{\zeta} = (\theta, \eta, \mu)$. The state space $\mathcal{S}$ of the process $Y(\boldsymbol{\zeta}) = \{Y(\boldsymbol{\zeta}, t), t \geq 0\}$ is given by $\mathcal{S} = \{0, 1, \ldots, n\}$. We assume that all exponential times are statistically independent of each other, hence it is readily verified that $Y(\boldsymbol{\zeta})$ is a continuous time Markov chain (see, e.g., Kulkarni [33]) whose transition diagram is given in Figure 3. Observe that this Markov chain consists of $n-1$ transient states (states $s=1$ to $s=n$) and one absorbing state ($s=0$) that corresponds to all the nodes of the network being active.

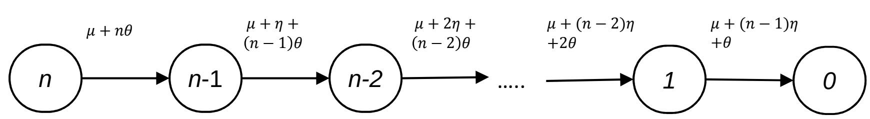


Figure 3 Transition diagram of the continuous time Markov chain $Y(\boldsymbol{\zeta})$. The number in each state is the number of inactive nodes in the network.

From the point of view of management and control of the process of nodes' activation, an important question is "which values of the parameters θ, η, and μ lead to the fastest activation of all nodes in the network?" It is easy to see, though, that increasing the values of the parameters θ, η, and μ leads to an increase in the probability that all nodes are activated by a given time t (see also Remark 1). In this example, we make an assumption that a unit increase in the rate η incurs a cost $c_\eta = c$; similarly, the rates μ and θ carry the costs $c_\mu = Kc$ and $c_\theta = Lc$, respectively, where $K, L > 0$, and there is a budget $B = 1$ available for covering these costs. Thus, the problem of optimizing the nodes' activation time becomes a resource allocation problem with the budget constraint

$$c(K\mu+\eta+L\theta)\leq 1.$$

REMARK 1. Recall that if a random variable A is exponentially distributed, $A\sim\exp(\alpha)$, then $E[A]=1/\alpha$ and $\mathrm{Var}[A]=1/\alpha^2$. Hence, if $A\sim\exp(\alpha)$, $B\sim\exp(\beta)$, and $\alpha>\beta$, then $E[A]<E[B]$ and $\mathrm{Var}[A]<\mathrm{Var}[B]$. In addition, $\alpha>\beta$ implies that A is less than B in the *usual stochastic order* ($A\preceq B$). That is, for all $t\geq 0$, $P[A\leq t]>P[B\leq t]$ (hence, is it more likely for A to take on smaller values than B). In particular, $A\preceq B$ is equivalent to $E[F(A)]\leq E[F(B)]$ for any non-decreasing function F, see Shaked and Shanthikumar [45]. ∎

As it will be shown in the next section, the problem of minimizing the activation times of all the nodes in the network can be framed as choosing the value of $\boldsymbol{\zeta}$ that minimizes the second-largest eigenvalue λ of the generator matrix of the Markov chain $Y(\boldsymbol{\zeta})$. In this particular example such a minimization problem, after further simplifications, reduces to a linear programming (LP) problem

$$\begin{aligned}\lambda^*=\min_{\lambda,\theta,\eta,\mu}\big\{\lambda:\ &\lambda+\mu+n\theta\geq 0,\ \lambda+\mu+(n-1)\eta+\theta\geq 0,\\ &c(K\mu+\eta+L\theta)\leq 1,\ \mu,\ \eta,\ \theta\geq 0\big\}.\end{aligned}$$

An optimal solution (θ^*,η^*,μ^*) of this LP problem and the corresponding optimal value λ^* can be computed analytically:

$$\lambda^*=-\frac{n}{c(L+1)}\quad\text{and}\quad\mu^*=0,\ \eta^*=\frac{1}{c(L+1)},\ \theta^*=\frac{1}{c(L+1)},\qquad\text{if } n\geq\frac{L+1}{K},\tag{1a}$$

$$\lambda^*=-\frac{1}{Kc}\quad\text{and}\quad\mu^*=\frac{1}{Kc},\ \eta^*=\theta^*=0,\qquad\text{if } n<\frac{L+1}{K}.\tag{1b}$$

This solution has several remarkable properties. First, it is the "either-or" form of the optimal resource allocation strategy, which states that it is *never* optimal to invest in all three activation modes: depending on the relative costs of the intrinsic capability rate θ, the external influence rate μ, and the size n of the network, one should either invest in improvement of the intrinsic capability rate θ and the network influence rate η (case (1a)), or the external influence rate μ (case (1b)). Second, solution (1a)–(1b) suggests that the larger the network is, the more likely it is optimal to invest in the network influence. And if this is the case, then λ^* decreases with n, which implies, as it will be seen below, that the probability of all nodes becoming active by a given time t increases to one exponentially fast with n, demonstrating the domino-type effect observed in cascade processes on networks.

3.2. Model Formulation

In this section we present a general model where the lower-level graphs contain more than two nodes (states). We consider an undirected complete network $G = (V, E)$, where each node $i \in V$ represents a particular *system* and an arc $(i, j) \in E$ between two nodes implies that nodes i and j can mutually influence each other. At any given time $t \in \mathbb{R}_+$, node i can be in a state $q \in \mathcal{Q} := \{1, 2, \ldots, Q+1\}$, where $\mathcal{Q}$ is assumed to be the same for each node $i \in V$. In the set $\mathcal{Q}$, state $q = 1$ represents the initial *inactive* state of a node, and state $q = Q + 1$ represents the desired *active* state. States $q = 2, \ldots, Q$ denote some intermediate levels of achieving the target active condition, where the degree of achievement increases with q. It is assumed that once a node reaches the active state $q = Q + 1$, it never leaves this state.

We assume that the *intrinsic capability* of a node is measured by a parameter $\theta \geq 0$, such that the time it takes a node to transition to state $Q+1$ from a current state $q < Q+1$ is exponentially distributed with the rate

$$g_q(h_q + \theta), \tag{2}$$

where $g_q, h_q \geq 0$, $q = 1, \ldots, Q$, are constants. Recall, by Remark 1, that an increase in the value of θ implies a "reduction" in the time until entering the active state.

The nodes that have transitioned to the active state may exert influence, through the network connections, on the remaining inactive or partially active nodes so that the latter improve their states. This *network influence* is quantified by non-negative parameters $\eta_1, \ldots, \eta_Q$, where η_k represents a node's capability to go from state q to state $q + k$ due to the influence/assistance by the active nodes. Formally, if a node is in state $q < Q + 1$, then the time that it takes for this node to transition to a better state r, $r > q$, due to the network influence from the active nodes is exponentially distributed with the rate

$$s_{Q+1} \frac{d_{q,r-q}(f_{q,r-q} + \eta_{r-q})}{s_q}, \tag{3}$$

where s_{Q+1} is the number of active nodes, s_q is the number of nodes in state q, and $d_{q,r-q}, f_{q,r-q}$ are nonnegative constants. The factor s_{Q+1} implies that the network influence effect strengthens as the number of active nodes increases, and s_q in the denominator of (3) means that the network influence is distributed uniformly to all nodes in the state q. As with θ, larger values of η_{r-q} imply faster transitions to improved states.

Finally, inactive and partially active nodes may be subject to *external influence* by an entity exogenous to the network whose function is to assist the nodes in becoming active. The capabilities of this external entity are modeled by nonnegative parameters $\mu_1, \ldots, \mu_Q$, where μ_k is the entity's capability to bring a node in the state q to the state $q + k$. The external entity's influence is distributed equally to all nodes in a given state, such that if a node is in state $q < Q+1$, then the time it takes this node to transition to a better state $r > q$ due to the external influence is exponentially distributed with the rate

$$\frac{a_{q,r-q}(b_{q,r-q} + \mu_{r-q})}{s_q}, \tag{4}$$

where $a_{q,r-q}$ and $b_{q,q-r}$ are nonnegative constants. As before, larger values of μ_{r-q} imply shorter times until activation.

Observe that in the general state model the vector of parameters is given by $\boldsymbol{\zeta} = (\theta, \eta_1, \ldots, \eta_Q, \mu_1, \ldots, \mu_Q)$. Moreover, hereafter, we make the assumption that all the exponential times defined above are statistically independent, and that the distributions of these times are the same for any node of the network.

3.2.1. Markov Chain Model. For a given value of $\boldsymbol{\zeta}$ let us define $Y(\boldsymbol{\zeta}, t)$ as the state of the network at time t, and let $Y(\boldsymbol{\zeta}) = \{Y(t, \boldsymbol{\zeta}) : t \geq 0\}$ be the associated stochastic process. As the distributions of the exponential times in Section 3.2 are the same for all nodes, it follows that the state space $\mathcal{S}$ of $Y(\boldsymbol{\zeta})$ can be completely characterized by

$$\mathcal{S} = \Big\{(s_1, \ldots, s_{Q+1}) : \sum_{q \in \mathcal{Q}} s_q = n\Big\},$$

where $n = |V|$, and s_q is the number of nodes in the network that are in the state q, $q \in \mathcal{Q}$. Observe that $\mathcal{S}$ contains the *absorbing state* $\boldsymbol{s}_A = (0, \ldots, 0, n)$, which corresponds to all the nodes of the network being active.

In order to compute the possible transitions between states in $\mathcal{S}$, note that at any given time any of the nodes in the state $q \neq Q+1$ can improve its activation state from q to r, $r > q$. That is, given that the process is in the state $\boldsymbol{s} \in \mathcal{S}$, it can jump to any state $\boldsymbol{s}'$ of the form

$$\boldsymbol{s}' = \boldsymbol{s} - \boldsymbol{e}_q + \boldsymbol{e}_r \qquad \text{for any } q, r \in \mathcal{Q}, \text{ with } q < r, \tag{5}$$

as long as $s_q > 0$. In equation (5), for any $q \in \mathcal{Q}$, $\boldsymbol{e}_q$ is a vector of size $Q+1$ with zeros everywhere except for a one in the q-th position. A jump from $\boldsymbol{s}$ to $\boldsymbol{s} - \boldsymbol{e}_q + \boldsymbol{e}_r$ can be triggered by a successful completion of either a network or a external influence event, and, if $r = Q+1$, it can also a result of a (completed) activation due to the intrinsic capability of a node. Since all the times associated with these events are exponentially distributed and independent, it can be shown by using standard Markov chains modeling arguments (see, e.g., Kulkarni [33]) that $Y(\boldsymbol{\zeta})$ is a CTMC whose transition rate between $\boldsymbol{s}$ and $\boldsymbol{s} - \boldsymbol{e}_q + \boldsymbol{e}_r$ is given by $\xi_{qr}(\boldsymbol{s}, \boldsymbol{\zeta})$, where

$$\xi_{qr}(\boldsymbol{s}, \boldsymbol{\zeta}) = \begin{cases} s_{Q+1} d_{q,j}(f_{q,j} + \eta_j) + a_{q,j}(b_{q,j} + \mu_j), & j = r - q, \text{ if } r < Q+1 \\ s_q g_q(h_q + \theta) + s_{Q+1} d_{q,j}(f_{q,j} + \eta_j) + a_{q,j}(b_{q,j} + \mu_j), & j = Q+1-q, \text{ if } r = Q+1. \end{cases}$$

Therefore, if we denote by $\Xi(\boldsymbol{\zeta})$ the *generator matrix* of $Y(\boldsymbol{\zeta})$, its off-diagonal entries are given by

$$(\boldsymbol{\Xi}(\boldsymbol{\zeta}))_{\boldsymbol{s}\boldsymbol{s}'} = \begin{cases} \xi_{qr}(\boldsymbol{s},\boldsymbol{\zeta}), & \text{if } \boldsymbol{s}' = \boldsymbol{s} - \boldsymbol{e}_q + \boldsymbol{e}_r \text{ for some } q, r \in \mathcal{Q},\ q < r \\ 0, & \text{otherwise,} \end{cases} \tag{6}$$

and its diagonal entries are given by

$$(\boldsymbol{\Xi}(\boldsymbol{\zeta}))_{\boldsymbol{s}\boldsymbol{s}} = -\sum_{\boldsymbol{s}' \in \mathcal{S}, \boldsymbol{s}' \neq \boldsymbol{s}} (\Xi(\boldsymbol{\zeta}))_{\boldsymbol{s}\boldsymbol{s}'}. \tag{7}$$

Particularly, note that $(\boldsymbol{\Xi}(\boldsymbol{\zeta}))_{\boldsymbol{s}\boldsymbol{s}'}$ can be interpreted as the rate of the exponentially distributed time that it takes $Y(\boldsymbol{\zeta})$ to jump from $\boldsymbol{s}$ to $\boldsymbol{s}'$. Similarly, $-\Xi(\boldsymbol{\zeta})_{\boldsymbol{s}\boldsymbol{s}}$ can be interpreted as the rate of the exponential time that $Y(\boldsymbol{\zeta})$ stays in the state $\boldsymbol{s}$ before jumping to other state, i.e., the *sojourn time* of process $Y(\boldsymbol{\zeta})$ in the state $\boldsymbol{s}$ (Kulkarni [33]).

3.2.2. Optimization Problem. Noting that all nodes are initially inactive, let $T(\boldsymbol{\zeta})$ be the time until all the nodes of the network become active. Observe that $T(\boldsymbol{\zeta})$ is precisely the time until the Markov chain $Y(\boldsymbol{\zeta})$ enters the absorbing state $\boldsymbol{s}_A = (0, \ldots, 0, n)$ given that at time $t = 0$ the chain is in the state $(n, 0, \ldots, 0)$. In this work, we are interested in "minimizing $T(\boldsymbol{\zeta})$," and, to this end, we pursue an approach based on maximizing the probability $P[T(\boldsymbol{\zeta}) \leq t]$ for $t \geq 0$.

The value of probability $P[T(\boldsymbol{\zeta}) \leq t]$ can be computed exactly or approximately using a number of analytical expressions that are given in terms of the generator matrix $\boldsymbol{\Xi}(\boldsymbol{\zeta})$ (Kulkarni [33]). In the optimization context, however, these expressions can be difficult to use due to their non-convexity in $\boldsymbol{\zeta}$. Instead, we pursue an approach based on optimizing a lower bound on $P[T(\boldsymbol{\zeta}) \leq t]$. In this regard, the following result holds.

Proposition 1. *Let $\boldsymbol{\zeta} > \mathbf{0}$ be given and consider the generator matrix $\boldsymbol{\Xi}(\boldsymbol{\zeta})$ of the Markov chain $Y(\boldsymbol{\zeta})$. Then, the largest eigenvalue of $\boldsymbol{\Xi}(\boldsymbol{\zeta})$ is zero, its second largest eigenvalue $\lambda_2(\boldsymbol{\zeta})$ is strictly negative, and there exist constants $M(\boldsymbol{\zeta}) \in \mathbb{R}$, $t_0(\boldsymbol{\zeta})$ such that*

$$P[T(\boldsymbol{\zeta}) \leq t] \geq 1 - e^{\lambda_2(\boldsymbol{\zeta})t} t^{K(\boldsymbol{\zeta})-1} M(\boldsymbol{\zeta}) \quad \forall t \geq t_0(\boldsymbol{\zeta}) \tag{8}$$

where $K(\boldsymbol{\zeta}) \geq 1$ is the algebraic multiplicity of $\lambda_2(\boldsymbol{\zeta})$.

Proposition 1 enables one to formulate the problem of "minimization" of $T(\boldsymbol{\zeta})$ via minimization of the second largest eigenvalue of $\boldsymbol{\Xi}(\boldsymbol{\zeta})$. The second largest eigenvalue minimization (SLEM) problem has been employed in other contexts involving optimization of the convergence rates of Markov chains, see, e.g., Sun et al. [46]. In the present framework, it is assumed that a unit increase in the value of a parameter $p \in \mathcal{P} = \{\theta, \eta_1, \ldots, \eta_Q, \mu_1, \ldots, \mu_Q\}$ incurs a cost of c_p, which leads to the following budget-constrained SLEM problem

$$\lambda^* = \min_{\boldsymbol{\zeta}} \Big\{ \lambda_2(\boldsymbol{\zeta}) : \sum_{j \in \mathcal{J}} \big(c_{\eta_j} \eta_j + c_{\mu_j} \mu_j \big) + c_\theta \theta \leq 1,\ \boldsymbol{\zeta} \geq \mathbf{0} \Big\}, \tag{9}$$

where $\mathcal{J} = \{1, \ldots, Q\}$. We refer to the above problem as *the cascade propagation optimization problem* and to λ^* as the network's *optimal cascade rate*. For general generator matrices, formulation (9) is computationally challenging, see Lewis and Overton [34]. But in the special case of the Markov chain discussed above, $\boldsymbol{\Xi}(\boldsymbol{\zeta})$ can be arranged as an upper-triangular matrix (see the proof of Proposition 1), and as such, its second largest eigenvalue is given by its largest nonzero diagonal element. Since the diagonal element of $\boldsymbol{\Xi}(\boldsymbol{\zeta})$ associated with the state $\boldsymbol{s}_A$ is zero, problem (9) can be equivalently posed as

$$\lambda^* = \min_{\boldsymbol{\zeta}} \left\{ \max_{\boldsymbol{s} \in \mathcal{S} \setminus \{\boldsymbol{s}_A\}} \{\Xi_{\boldsymbol{ss}}(\boldsymbol{\zeta})\} : \sum_{j \in \mathcal{J}} \big(c_{\eta_j}\eta_j + c_{\mu_j}\mu_j\big) + c_\theta \theta \leq 1, \ \boldsymbol{\zeta} \geq \mathbf{0} \right\}. \tag{10}$$

By using standard LP techniques, and by recalling the definition of $\Xi_{\boldsymbol{ss}}(\boldsymbol{\zeta})$ in equations (6) and (7), problem (10) can be transformed into the following LP problem

$$\begin{aligned} \lambda^* = \min \quad & \lambda \\ \text{s. t.} \quad & \lambda + \sum_{j \in \mathcal{J}} A_{\boldsymbol{s}j}\mu_j + \sum_{j \in \mathcal{J}} D_{\boldsymbol{s}j}\eta_j + G_{\boldsymbol{s}}\theta \geq -K_{\boldsymbol{s}}, \quad \forall \boldsymbol{s} \in \widetilde{\mathcal{S}} \\ & \sum_{j \in \mathcal{J}} \big(c_{\eta_j}\eta_j + c_{\mu_j}\mu_j\big) + c_\theta\theta \leq 1 \\ & \mu_j, \ \eta_j \geq 0, \ j \in \mathcal{J}, \ \theta \geq 0, \ \lambda \in \mathbb{R}. \end{aligned} \tag{11}$$

In the above program, $\widetilde{\mathcal{S}} := \mathcal{S} \setminus \{\boldsymbol{s}_A\}$, the constants $A_{\boldsymbol{s}j}$, $D_{\boldsymbol{s}j}$ and $G_{\boldsymbol{s}}$ are defined by

$$A_{\boldsymbol{s}j} := \sum_{q=1}^{Q-j+1} \mathbb{1}_{\{s_q > 0\}} a_{qj}, \ D_{\boldsymbol{s}j} := \sum_{q=1}^{Q-j+1} \mathbb{1}_{\{s_q > 0\}} s_{Q+1} d_{qj}, \ G_{\boldsymbol{s}} := \sum_{q=1}^{Q} s_q g_q,$$

for all $j \in \mathcal{J}$ and $s \in \widetilde{\mathcal{S}}$, with $\mathbb{1}_A$ being the indicator function of set A. Also, $K_{\boldsymbol{s}}$ is given by

$$K_{\boldsymbol{s}} := \sum_{j \in \mathcal{J}} \left(\sum_{q=1}^{Q-j+1} \mathbb{1}_{\{s_q > 0\}} (a_{qj} b_{qj} + s_{Q+1} d_{qj} f_{qj}) \right) + \sum_{q=1}^{Q} s_q g_q h_q, \quad \forall \boldsymbol{s} \in \widetilde{\mathcal{S}}.$$

In the next section we derive, under some reasonable assumptions, certain properties of the proposed model that lead to a simple analytical solution of (11). Furthermore, the obtained results provide valuable insights into the structure of the structure of the optimal cascade propagation in complete networks.

3.3. Optimal solution, Analysis, and Discussion

In this section, and hereafter, we make the following assumptions regarding the costs and parameters of the model in complete networks:

(**A1**) Costs c_{μ_j} and c_{η_j} are non-decreasing in j.

(**A2**) Parameters a_{qj} and d_{qj} are non-increasing in j for any $q \in \mathcal{Q} \setminus \{Q+1\}$.

Assumption (**A1**) reflects the intuitive fact that *"longer jumps are costlier,"* where by a "long jump" we mean a transition from a state q to state $q+k$, $k \geq 2$, during which a node bypasses several intermediate stages of the activation process. Assumption (**A2**) states that the likelihood of a successful partial progress towards activation is non-increasing with respect to the influence effort, or, loosely speaking, *"longer jumps are less likely than short steps."*

PROPOSITION 2. *Suppose that assumptions* **(A1)** *and* **(A2)** *hold. Then, there exists an optimal solution of problem* (11) *that satisfies* $\mu_j = 0$ *and* $\eta_j = 0$ *for all* $j > 1$.

REMARK 2. This result implies that *"short steps can be as optimal as long leaps."* More precisely, under conditions (**A1**)–(**A2**), an "incremental," or "step-by-step" influence process can provide the same guarantee of a fast cascade rate for the entire network as an "aggressive," or "accelerated" process during which nodes may skip several intermediate activation stages. Consequently, if it is optimal to provide network or external influence to the nodes, then *it is not necessary to influence the nodes in a "long-jump" fashion*. The intuition behind this result is that single-jump influence is useful at every state of the cascade of a node. Indeed, observe that any other influence effort cannot be used at all levels (e.g., if a node is at state Q, it is not necessary to provide it with influence that makes its level jump by two or more states).

REMARK 3. In view of Proposition 2, from now on *we assume that an optimal solution of* (11) *always satisfies* $\mu_j = 0$ and $\eta_j = 0$ for $j > 1$. Thus, for convenience we set $\mu := \mu_1$ and $\eta := \eta_1$. Moreover, we refer to μ, η, and θ as the *external influence*, the *network influence* and the *intrinsic capability* rates, respectively.

In what follows, to simplify the notation, let $a := a_{11}$, $d := d_{11}$, and $g := g_1$. Observe that a and d measure, respectively, how effective the external and network influence are in assisting a node to improve its state from $q = 1$ (inactive) to $q = 2$. Likewise, g measures how effective is the intrinsic capability when a node is inactive. We introduce the following assumption:

(**A3**) Inequalities $a \leq a_{q1}$, $d \leq d_{q1}$, and $g \leq g_q$ hold for all $q = 2, \ldots, Q$.

This assumption implies that *"the first step toward activation is the most difficult,"* or, it is more likely for network or external influence or intrinsic capability to be successful when the node is in a partially activated state ($q > 1$) than when the node is inactive ($q = 0$).

Finally, define τ as

$$\tau := \frac{nc_\theta d}{c_\theta d + c_\eta g},$$

It is easily verified that τ is the point where gx/c_θ and $d(n-x)/c_\eta$, as functions of x, coincide, i.e.,

$$\frac{g\tau}{c_\theta} = \frac{d(n-\tau)}{c_\eta} = \frac{dgn}{c_\theta d + c_\eta g}.$$

Proposition 3. *Suppose that assumptions* **(A1)**, **(A2)**, *and* **(A3)** *hold, and that the rate functions $\xi(\boldsymbol{s},\boldsymbol{\zeta})$ are linear, i.e., the constants h_q, b_{qj} and f_{qj} take the value zero for all $q,j\in\mathcal{J}$. Then, only the constraints associated with the states $(n,0,\ldots,0)$ and $(1,0,\ldots,n-1)$ have to be included into LP (11). That is, problem (11) reduces to*

$$\min_{\lambda,\theta,\eta,\mu}\big\{\lambda:\ \lambda+a\mu+ng\theta\geq 0,\ \ \lambda+a\mu+(n-1)d\eta+g\theta\geq 0,\\ c_\mu\mu+c_\eta\eta+c_\theta\theta\leq 1,\ \ \mu,\ \eta,\ \theta\geq 0\big\},\tag{12}$$

and its optimal solution is given in Table 1:

Case	λ^*	μ^*	η^*	θ^*
$\tau\geq 1,\ \frac{g\tau}{c_\theta}\leq\frac{a}{c_\mu}$	$-\frac{a}{c_\mu}$	$\frac{1}{c_\mu}$	0	0
$\tau\leq 1,\ \frac{g}{c_\theta}\leq\frac{a}{c_\mu}$	$-\frac{a}{c_\mu}$	$\frac{1}{c_\mu}$	0	0
$\tau\geq 1,\ \frac{a}{c_\mu}\leq\frac{g\tau}{c_\theta}$	$-\frac{g\tau}{c_\theta}$	0	$\frac{g}{c_\theta d+c_\eta g}$	$\frac{d}{c_\theta d+c_\eta g}$
$\tau\leq 1,\ \frac{a}{c_\mu}\leq\frac{g}{c_\theta}$	$-\frac{g}{c_\theta}$	0	0	$\frac{1}{c_\theta}$

Table 1 An optimal solution of program (11) under assumptions (**A1**), (**A2**), (**A3**), and linearity of the rate functions.

Further analysis of Proposition 3 reveals several interesting properties of an optimal solution of problem (12) presented in Table 1:

Bottleneck states of the cascade process: The reduction of program (11) to (12) is a consequence of the fact that for all $\boldsymbol{s}\in\widetilde{\mathcal{S}}$ the negative diagonal elements of the generator matrix $-\boldsymbol{\Xi}_{\boldsymbol{ss}}(\boldsymbol{\zeta})$ are bounded from below by a convex combination of $-\boldsymbol{\Xi}_{\boldsymbol{s}_1\boldsymbol{s}_1}(\boldsymbol{\zeta})$ and $-\boldsymbol{\Xi}_{\boldsymbol{s}_n\boldsymbol{s}_n}(\boldsymbol{\zeta})$, where $\boldsymbol{s}_1=(1,0,\ldots,n-1)$ and $\boldsymbol{s}_n=(n,0,\ldots,0)$. Since $-\boldsymbol{\Xi}_{\boldsymbol{ss}}(\boldsymbol{\zeta})$ is the rate of the exponentially distributed time that $Y(\boldsymbol{\zeta})$ stays in $\boldsymbol{s}$, this result implies that the time that process $Y(\boldsymbol{\zeta})$ spends in any $\boldsymbol{s}\in\widetilde{\mathcal{S}}$ is stochastically smaller (see Remark 1) than either the time it spends on $\boldsymbol{s}_1$ or the time it spends on $\boldsymbol{s}_n$. In this sense, the smaller is the value of $-\boldsymbol{\Xi}_{\boldsymbol{ss}}(\boldsymbol{\zeta})$, the longer is the time the process spends in $\boldsymbol{s}$ (and hence, the slower the absorbtion time). Therefore, states $\boldsymbol{s}_1$ and $\boldsymbol{s}_n$ can be viewed as the *bottlenecks* of the cascade process.

The bottleneck states $\boldsymbol{s}_1$ and $\boldsymbol{s}_n$ also represent the "extreme" states of the cascade from the resource allocation standpoint. On one hand, in the state $\boldsymbol{s}_1$ all nodes but one are active, and hence in this state the cascade observes the maximum effect of any investment into network influence (because the active nodes assist in activation of the remaining node) and the minimum effect of the intrinsic capability. On the other hand, in the state $\boldsymbol{s}_n$ all nodes are inactive, and thus in this state the cascade observes the maximum effect of any investment into an intrinsic capability and the least effect of any investment into network influence (in fact, the latter is zero by its definition as all nodes are inactive i.e., $s_{Q+1}=0$).

Optimal solution structure: In addition, the optimal solution established by Proposition 3 has the remarkable property that it is never optimal to invest in all three activation modes simultaneously and it is never optimal to invest only into the network influence rate η. Particularly, if it is optimal to invest in the network influence rate η, then it is necessary to invest in the intrinsic capability rate θ as well (recall also our discussion of the bottleneck property of s_n). Furthermore, under no circumstances both the network influence rate η and the external influence rate μ may be both non-zero. And, if it is optimal to provide the external influence, then all resources must be invested into the external influence.

Network interactions and cascade propagation: Importantly, Proposition 3 reveals the *significance of the network interactions for the speed of the cascade propagation.* Whenever it is optimal to exploit the interactions between nodes, i.e., whenever $\eta^* > 0$, the rate λ^* decreases linearly with the size of the network n. It implies, by virtue of (8), that for any time $t > 0$ *the probability that all the nodes are active by time t increases exponentially to one with n* (see Section 3.4 for a more detailed discussion of this fact). In contrast, and as expected, if in the optimal solution the network interactions do not play a role (i.e., $\eta^* = 0$), then λ^* does not depend on n and thus the cascade propagation is independent of the size of the network.

Observe that the conditions in the third row of Table 1 can be restated as $d(n-1)/c_\eta \geq g/c_\theta$ and $d(n-\tau)/c_\eta \geq a/c_\mu$. In other words, if the network is sufficiently large or the cost-adjusted network influence (d/c_η) is relatively large with respect to the cost-adjusted rates of the other two modes (g/c_θ and a/c_μ), then it is optimal to exploit network interactions. These conditions are naturally satisfied in many applications. For example, in social networks and online social networks, word of mouth/viral marketing is generally effective and cheap to implement, see e.g., Buttle [12], Aral and Walker [4], Bapna and Umyarov [7], and these networks are likely to be large.

These conclusions suggest that a more formal analysis of the sensitivity of λ^* with respect to n should take into account that the budget and the costs c_μ, c_η and c_θ might depend on n as well, particularly if n is large. We conduct this analysis in the next section.

3.4. Structure of Optimal Solutions for Large-Scale Networks

In this section we study the behavior of optimal solutions when the number of nodes in the network is very large, under the assumption that the cost parameters and the budget are functions of n, i.e., the budget constraint of program (11) has the form $c_\mu(n)\mu + c_\eta(n)\eta + c_\theta(n)\theta \leq B(n)$. To simplify the discussion, we set $c'_p(n) = c_p(n)/B(n)$ for all parameters $p \in \mathcal{P}$, and refer to $c'_p(n)$ as the *adjusted cost* of increasing the parameter p.

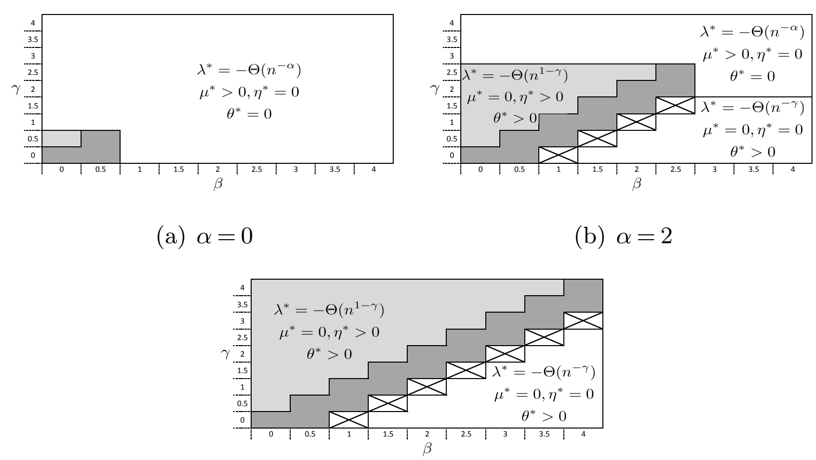
Furthermore, for the sake of simplicity, we make the assumption that the adjusted costs are asymptotic powers of n. A direct application of Proposition 3 yields the following result:

Corollary 1. *Suppose that $c'_\mu(n) = \Theta(n^\alpha)$, $c'_\eta(n) = \Theta(n^\beta)$ and $c'_\theta(n) = \Theta(n^\gamma)$, where $\alpha, \beta, \gamma \geq 0$. Then, the asymptotic behavior of optimal solutions of program (11) is given by Table 2. Furthermore, in any case not included in the Cases of the left column of Table 2 (i.e., when there are strict equalities), optimal solutions depend on the values of the constants a, d and g.*

Case	λ^*	μ^*	η^*	θ^*
(A)	$-\Theta(n^{-\alpha})$	$\Theta(n^{-\alpha})$	0	0
(B)	0	0	$\Theta(n^{-\gamma})$	
(C)	$-\Theta\big(n^{1-\max\{\beta,\gamma\}}\big)$	0	$\Theta\big(n^{-\max\{\beta,\gamma\}}\big)$	$\Theta\big(n^{-\max\{\beta,\gamma\}}\big)$

Table 2 An asymptotic behavior of optimal solutions of program (11) under conditions of Corollary 1. Case (A): $\beta < \gamma + 1$ and ($\alpha + 1 < \gamma$ or $\alpha + 1 < \beta$) or ($\gamma + 1 < \beta$ and $\alpha \leq \gamma$); Case (B): $\gamma + 1 < \beta$ and $\gamma < \alpha$; Case (C): $\beta < \gamma + 1$ and $\gamma < \alpha + 1$ and $\beta < \alpha + 1$.

The corollary becomes more informative if shown graphically. To this end, we illustrate it in the case where β and γ are multiples of 1/2 and $0 \leq \alpha, \beta, \gamma \leq 4$, see Figure 4.



(a) $\alpha = 0$ (b) $\alpha = 2$

(c) $\alpha = 4$

Figure 4 Graphical representation of Corollary 1. The light gray and gray regions correspond to $\lambda^* = \Theta(n^{1-\gamma})$ and $\lambda^* = \Theta(n^{1-\beta})$, respectively. In both cases $\mu^* = 0$, $\eta^*, \theta^* > 0$. The squares with an "$\times$" represent. that for these particular values of β and γ the optimal solution is given by one of the neighboring regions depending on the values of constant parameters a, d, and g.

In Figure 4(a) observe that if $\alpha = 0$, i.e., the external influence cost has the same asymptotic behavior as the budget, then it is optimal for most cases to invest only into the external influence. This yields an optimal cascade rate λ^* that is independent of the size of the network. However, if both the adjusted network influence cost, $c'_\eta(n)$, and the adjusted intrinsic capability cost, $c'_\theta(n)$, are sufficiently small (at most of the order of $n^{1/2}$, so it is relatively cheap to invest into them), then the optimal solution is to share the investments between the network influence and intrinsic capability. Moreover, the optimal cascade rate becomes a strictly decreasing function of n.

Consider, on the other hand, Figure 4(c), where $\alpha = 4$ (the analysis of Figure 4(b) is similar). Note that, as expected, the optimal solution does not involve investing into the external influence. On the other hand, depending on the relationship between β and γ, there are two different optimal budget allocations. The first optimal allocation corresponds to the case of $\gamma + 1 < \beta$ (white region), i.e., when the per-node investment into the network influence is significantly larger than the investment into intrinsic capability. Then, the optimal solution is to invest only into the intrinsic capability, and λ^* is a non-decreasing (strictly decreasing if $\gamma > 0$) function of n. In the second allocation, when $\beta \leq \gamma + 1$, it is optimal to invest in both the network influence and intrinsic capability and the optimal cascade rate is $-\Theta(n^{1-\gamma})$ if $\beta < \gamma$ (light gray), else it is $-\Theta(n^{1-\beta})$ (dark gray). In particular, if $c'_\eta(n)$ and $c'_\theta(n)$ are at most of the order of $n^{1/2}$, then λ^* is again a strictly decreasing function of n; otherwise it is nondecreasing.

An important conclusion of Figure 4 (and more generally of Corollary 1) is that the size of the network has an opposite effect on the rate at which the cascade spreads, see Figure 5, where we plot t versus $1 - e^{-\lambda^* t}$, the absorbtion probability lower-bound given by Proposition 1 (assuming that $M(\boldsymbol{\zeta}) = K(\boldsymbol{\zeta}) = 1$). Indeed, if the orders of both ratios $c_\eta(n)/B(n)$ and $c_\theta(n)/B(n)$ are strictly less than n, then λ^* is strictly decreasing in n, which implies that *the larger is the number of nodes in the network, the faster the cascade spreads*; see Figure 5 (c). This highly desirable behavior is, as discussed in Section 3.3, a consequence of the network effect (i.e., $\eta^* > 0$), and is satisfied in settings where economies of scale are observed in the investments into the network influence and intrinsic capability. In contrast, if either $c_\eta(n)$ or $c_\theta(n)$ asymptotically increase with n at least as fast as $nB(n)$, then λ^* is nondecreasing in n and hence *the larger is the number of nodes in the network, the slower the cascade spreads*; see Figure 5 (a).

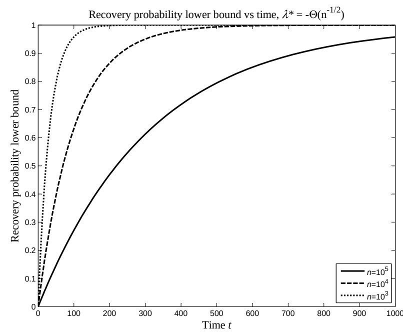
4. A Cascading Process in General Networks

Below we extend the model of Section 3 to networks with general topologies. In Section 4.1 we discuss the multiscale decomposition of the network, which leads to the introduction of the cascade propagation model in Section 4.2. The solution of the model, the relationship of the problem with a MSA on an auxiliary graph, and the hierarchical structure of the optimal solution, are discussed in Sections 4.3 and 4.4.

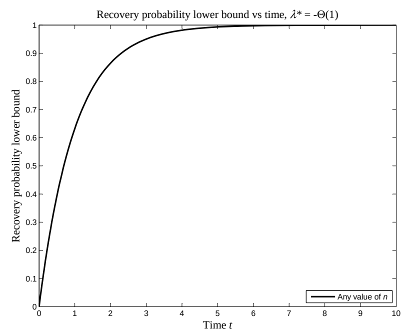
4.1. A Multiscale Decomposition of General Networks

Consider an undirected network $G = (V, E)$, and let its node set V be partitioned as

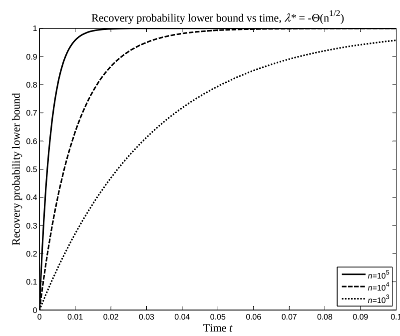
$$V = \bigcup_{k \in \mathcal{C}} V^k, \tag{13}$$



(a) λ^* increasing in n



(b) λ^* constant in n



(c) λ^* decreasing in n

Figure 5 Convergence of the network absorbtion probability lower bound to one as a function of time for different exponents of λ^* and values of n; see discussion in Section 3.4. Observe the magnitude of differences in the time axis between the different cases.

where $\mathcal{C} := \{1,\dots,N\}$ for $N \in \mathbb{Z}_{>0}$. We denote by $C^k := G[V^k]$ the graph induced by V^k in G. The collection of all these graphs comprises the *medium-level graphs* of the multi-scale decomposition of G, see Figure 1(b).

We assume that the partition in (13) is such that each C^k is a complete subgraph, or a clique, of G. This assumption is motivated by the fact that a clique is an ideal representation of a tightly-knit cluster of nodes with similar properties. Indeed, nodes in a clique have the best possible level of familiarity (all nodes are connected), reachability (there is only one arc between any pair of nodes), and robustness (cliques have maximum possible connectivity), see Balasundaram et al. [6]. As such, nodes in a clique are particularly susceptible to be be directly influenced by other nodes in the clique (Wasserman and Faust [48]).

The *high-level graph* is defined as a directed weighted graph $G^H := (\mathcal{C}, E^H)$. The weight $w^{\ell k}$ of the edge $(\ell, k) \in E^H$ measures the influence that the clique C^ℓ has on C^k, where $k \neq \ell$, and we adopt the convention that there is a directed arc in E^H between k and ℓ as long as $w^{\ell k} > 0$. In general, the weights $w^{\ell k}$ can have any arbitrary nonnegative value. However, since they measure the degree of influence between different cliques, we view them as functions of the structure of G.

Therefore, depending on the application, different approaches can be used to determine the value of $w^{\ell k}$, with two examples provided below:

(i) *The average number of arcs between cliques per node:* Let $E^{\ell k} \subseteq E$ be the edges in G between nodes in C^ℓ and C^k, that is, $E^{\ell k} = \{(i,j) \in E\colon (i,j) \in V^\ell \times V^k\}$. Then

$$w^{\ell k} = \frac{|E^{\ell k}|}{n^\ell n^k}, \tag{14}$$

where $n^k := |V^k|$ for any $k \in \mathcal{C}$. The intuition behind the above definition is that the greater is the number of arcs per node between cliques, the higher is their degree of influence. Particularly, cliques with no arcs between them cannot influence each other.

(ii) *The reciprocal of the average distance between nodes in cliques*: For any $(i,j) \in V^\ell \times V^k$ let p_{ij} be the number of arcs in the shortest path between i and j on G (also referred as the distance between i and j, see e.g., West et al. [51]). Then let

$$w^{\ell k} = \frac{n^\ell n^k}{\sum_{(i,j)\in V^\ell \times V^k} p_{ij}},$$

where we assume that $w^{\ell k} = 0$ if there are no arcs between the nodes in C^ℓ and C^k, i.e., $p_{ij} = +\infty$ for $i \in V^k$ and $j \in V^\ell$. This definition implies that the farther away the cliques are from each other, the less influence they exert on each other.

Any node $i \in V^k$ constitutes a *lower-level* graph that describes the possible stochastic transitions between the cascade states of that node. Following Section 3, we assume that at any given time a node may be in a state from the set $\mathcal{Q}$. The next section discusses the model of stochastic processes governing the transitions at lower-scale graphs, which generalizes the model presented in Section 3.2 and accounts for the influence that cliques exert on each other.

4.2. Model Formulation

Given a clique C^k, we assume that the cascade evolution within the clique follows the model developed in Section 3.2, with the exception that inactive or partially active nodes in C^k may be subject to network influence by not only the active nodes in C^k, but also by the active nodes in other cliques C^ℓ, $\ell \neq k$. The intrinsic capability, network influence, and external influence modes of activation are governed by the corresponding sets of parameters $\{\theta^k\}$, $\{\eta_j^{\ell k}\}$, and $\{\mu_j^k\}$, whose values must satisfy a resource allocation constraint.

Namely, let the set $\mathcal{Q}$ be defined as in Section 3. If a node in C^k is in a state $q \neq Q+1$, then the time it takes for a node to transition to the state $Q+1$ on its own is exponentially distributed with the rate (compare with (2))

$$g_q^k(h_q^k + \theta^k),$$

where g_q^k, h_q^k are nonnegative constants. Similarly, if a node in C^k is in a state $q \neq Q+1$, then the time it takes this node to transition to a state $r > q$ due to the external influence is exponentially distributed with the rate (compare with (4))

$$\frac{a_{q,r-q}^k(b_{q,r-q}^k + \mu_{r-q}^k)}{s_q^k},$$

where s_q^k is the number of nodes in the state q in the clique C^k, and $a_{q,r-q}^k, b_q^k \geq 0$ are constants. If a node in C^k is in a state $q \neq Q+1$, then its transition to a state $r > q$ can also be achieved due to the network influence by the active nodes in its own clique as well as the active nodes in any clique C^ℓ such that $w^{\ell k} > 0$, i.e., such that there is an arc in the high-scale graph between C^ℓ and C^k. The corresponding time is exponentially distributed with the rate (compare with (3))

$$w^{\ell k}\frac{s_{Q+1}^\ell d_{q,r-q}^{\ell k}(f_{q,r-q}^{\ell k} + \eta_{r-q}^{\ell k})}{s_q^k}, \tag{15}$$

where $d_{q,r-q}^{\ell k}, f_{q,r-q}^{\ell k}$ are nonnegative constants, s_{Q+1}^ℓ is the number of active nodes in the clique C^ℓ, and the factor $w^{\ell k}$ measures the degree of influence between the cliques C^ℓ and C^k in the high-level graph $\mathcal{G}^H$ (see above). When $\ell = k$ in (15), we set $w^{\ell k} = 1$. In the remainder of the paper, in order to simplify notation, and without loss of generality, we assume that $w^{\ell k}$ is incorporated into the value of the parameter $d_{r,r-q}^{\ell k}$.

Observe that this model of network influence defines two types of influence relationship between adjacent nodes in the network G. On one hand, there is a "Close" influence relationship between the adjacent nodes within the same clique, as equation (15) implies that the network influence rate of a node depends "directly" on the state of the nodes in its own clique. On the other hand, if the two adjacent nodes belong to different cliques, expression (15) implies that their influence relationship is "Distant," as the influence they exert on each other depends only "indirectly" on their state through the number of active nodes in each clique. That is, the influence that a node exerts on its non-clique neighbors is the same whether the node is active or inactive, as long as the number of active nodes in its clique remains the same.

This model of network influence allows a node to be influenced by non-adjacent nodes in the neighboring cliques, and, more importantly, allows for the network influence to depend on the sizes of the cliques. As such, per equation (15), larger cliques are a priori more influential than smaller cliques. We observe, however, that our model accommodates standard models of influence such as the ones discussed in, e.g., Kempe et al. [30], where the state of the cascade of a node depends "directly" on the state of all its neighbors and there is no "indirect" influence. Indeed, such a behavior can be achieved in our model by assuming that each node of the network G is a clique on its own.

In the multi-clique model the vector of the control parameters of the cascade is given by

$$\boldsymbol{\zeta} := \Big(((\mu_j^k)_{j\in\mathcal{J}}, \theta^k)_{k\in\mathcal{C}}, (\eta_j^{\ell k})_{j\in\mathcal{J},\ell,k\in\mathcal{C}} \Big).$$

Finally, we note that we are making the implicit assumption that the distribution of the above random times is the same for all nodes within a clique. Moreover, we assume that, given the values of s_q^k for all $q \in \mathcal{Q}$ and $k \in \mathcal{C}$, as well as the values of the elements of $\boldsymbol{\zeta}$, all the exponential random variables are statistically independent.

4.2.1. Markov Chain Model. Let $Y(t,\boldsymbol{\zeta})$ and $Y(\boldsymbol{\zeta})$ be defined as in Section 3.2.1. Since for any pair of nodes in the same clique the distributions of the exponential times discussed in Section 4.2 are the same, we can completely characterize the state space of $Y(\boldsymbol{\zeta})$ by

$$\mathcal{S} := \Big\{ (\boldsymbol{s}^1,\dots,\boldsymbol{s}^N) \in \prod_{k\in\mathcal{C}} \{0,\dots,n^k\}^{Q+1} : \ \boldsymbol{s}^k = (s_1^k,\dots,s_{Q+1}^k),\ \sum_{q\in\mathcal{Q}} s_q^k = n^k \quad \forall k \in \mathcal{C} \Big\},$$

where s_q^k is the number of nodes in the clique C^k in the state q and $n^k = |V^k|$. In particular, $\boldsymbol{s}_A = (0,\dots,n^1,0,\dots,n^2,\dots,0,\dots,n^N)$ is the absorbing state of $Y(\boldsymbol{\zeta})$, such that all the nodes in the network are active. Additionally, we note that the cardinality of $\mathcal{S}$ is given by

$$|\mathcal{S}| = \prod_{k=1}^{N} \binom{n^k+Q}{n^k}.$$

The transitions between the states in $\mathcal{S}$ are modeled similarly to those in Section 3.2.1. From any state $\boldsymbol{s}\in\mathcal{S}$ the process can only jump to states of the form $\boldsymbol{s} - \boldsymbol{e}_q^k + \boldsymbol{e}_r^k$, with $q,r\in\mathcal{Q}$, $q<r$, $k\in\mathcal{C}$, where for any r and k, $\boldsymbol{e}_r^k$ is a unit vector of dimension $N\cdot(Q+1)$ with a one in the $\big((k-1)(Q+1)+r\big)$-th position. Observe that the jump from $\boldsymbol{s}$ to $\boldsymbol{s}-\boldsymbol{e}_q^k+\boldsymbol{e}_r^k$ means that one of the nodes in the clique C^k moved from state q to state r. An analysis similar to that of Section 3.2.1 implies that $Y(\boldsymbol{\zeta})$ is a CTMC, whose transition rate between $\boldsymbol{s}$ and $\boldsymbol{s}-\boldsymbol{e}_q^k+\boldsymbol{e}_r^k$ is given by $\xi_{qr}^k(\boldsymbol{s},\boldsymbol{\zeta})$, where

$$\xi_{qr}^k(\boldsymbol{s},\boldsymbol{\zeta}) = \begin{cases} a_{q,j}^k(b_{q,j}^k+\mu_j^k) + \sum_{\ell\in\mathcal{C}} s_{Q+1}^\ell d_{q,j}^{\ell k}(f_{q,j}^{\ell k}+\eta_j^{\ell k}), & j=r-q,\ q<r,\ r\neq Q+1 \\ s_q^k g_q^k(h_q^k+\theta^k) + a_{q,j}^k(b_{q,j}^k+\mu_j^k) + \sum_{\ell\in\mathcal{C}} s_{Q+1}^\ell d_{q,j}^{\ell k}(f_{q,j}^{\ell k}+\eta_j^{\ell k}), & j=Q+1-q,\ q\leq Q. \end{cases}$$

It can be concluded that the off-diagonal elements of the generator matrix $\Xi_{\boldsymbol{s}\boldsymbol{s}'}(\boldsymbol{\zeta})$ are given by

$$\Xi_{\boldsymbol{s}\boldsymbol{s}'}(\boldsymbol{\zeta}) = \begin{cases} \xi_{qr}^k(\boldsymbol{s},\boldsymbol{\zeta}), & \text{if } \boldsymbol{s}' = \boldsymbol{s}-\boldsymbol{e}_q^k+\boldsymbol{e}_r^k,\ r\neq q,\ k\in\mathcal{C} \\ 0, & \text{otherwise}, \end{cases}$$

while its diagonal elements equal the negative sums of the corresponding rows, see equation (7).

4.2.2. Optimization Problem. As in the case of complete networks, the goal is to select the value of $\boldsymbol{\zeta}$ such that the distribution function $P[T(\boldsymbol{\zeta}) \leq t]$ approaches unity as early as possible. The values of the vector of parameters $\boldsymbol{\zeta}$ must satisfy the budget constraint

$$\sum_{k\in\mathcal{C}}\Big\{c_{\theta^k}\theta^k + \sum_{j\in\mathcal{J}}\big(c_{\mu_j^k}\mu_j^k + \sum_{\ell\in\mathcal{C}} c_{\eta_j^{\ell k}}\eta_j^{\ell k}\big)\Big\} \leq 1, \tag{16}$$

which is a straightforward generalization of the budget constraint for the complete network setting. Since Proposition 1 also holds for the multi-clique Markov chain, the problem of optimizing the Markov chain's convergence to the absorbing state can similarly be formulated in terms of minimizing the second largest eigenvalue of $\Xi(\boldsymbol{\zeta})$, see (9). In this case the generator matrix $\Xi(\boldsymbol{\zeta})$ can also be made upper-triangular (see the proof of Proposition 1), whereby the optimization of the second largest eigenvalue can be reduced to the following LP:

$$\begin{aligned} \lambda^* = \min \quad & \lambda \\ \text{s. t.} \quad & \lambda + \sum_{k\in\mathcal{C}}\Big\{G_{\boldsymbol{s}}^k\theta^k + \sum_{j\in\mathcal{J}}\big(A_{\boldsymbol{s}j}^k\mu_j^k + \sum_{\ell\in\mathcal{C}} s_{Q+1}^\ell D_{\boldsymbol{s}j}^{\ell k}\eta_j^{\ell k}\big)\Big\} \geq -K_{\boldsymbol{s}} \quad \forall \boldsymbol{s}\in\widetilde{\mathcal{S}} \\ & \text{constraint (16)},\ \theta^k,\ \mu_j^k,\ \eta_j^{\ell k} \geq 0,\ j\in\mathcal{J},\ \ell,k\in\mathcal{C},\ \lambda\in\mathbb{R}. \end{aligned} \tag{17}$$

As before, $\widetilde{\mathcal{S}} := \mathcal{S}\setminus\{s_A\}$ and for any $s\in\widetilde{\mathcal{S}}$, $k\in\mathcal{C}$ and $j\in\mathcal{J}$

$$G_{\boldsymbol{s}}^k = \sum_{q=1}^{Q} g_q^k s_q^k, \quad A_{\boldsymbol{s}j}^k = \sum_{q=1}^{Q-j+1} \mathbb{1}_{\{s_q^k>0\}} a_{qj}^k, \quad D_{\boldsymbol{s}j}^{\ell k} = \sum_{q=1}^{Q-j+1} \mathbb{1}_{\{s_q^k>0\}} d_{qj}^{\ell k},$$

and

$$K_{\boldsymbol{s}} = \sum_{k\in\mathcal{C}}\Big\{\sum_{q=1}^{Q}\Big(s_q^k h_q^k g_q^k + \mathbb{1}_{\{s_q^k>0\}}\Big[\sum_{r=q+1}^{Q} a_{q,r-q}^k b_{q,r-q}^k + \sum_{\ell\in\mathcal{C}} s_{Q+1}^\ell d_{q,r-q}^{\ell k} f_{q,r-q}^{\ell k}\Big]\Big)\Big\}.$$

Note that problem (17) has $N(QN+Q+1)$ variables and $\prod_{k=1}^N \binom{n^k+Q-1}{Q}$ constraints, which makes it prohibitively large to solve using standard software even for small networks. In the next sections, we discuss the conditions on the costs and parameters of this linear program that allow us to construct its optimal solution analytically. Moreover, our analysis provides important managerial insights into the structure of optimal cascade propagation in general networks and the corresponding resource allocation.

4.3. Basic Properties of Optimal Solutions

Next, we make the following assumptions, which generalize assumptions (**A1**)–(**A3**) of the single-clique case:

(**A1E**) Costs $c_{\mu_j}^k$ and $c_{\eta_j}^{\ell k}$ are non-decreasing in j for all $k,\ell\in\mathcal{C}$.

(**A2E**) Parameters a_{qj}^k and $d_{qj}^{\ell k}$ are non-increasing in j for any $q\in\mathcal{Q}\setminus\{Q+1\}$ and all $k,\ell\in\mathcal{C}$.

(**A3E**) It is satisfied that $a_{11}^k \le a_{q1}^k$, $d_{11}^{\ell k} \le d_{q1}^{\ell k}$, and $g_1^k \le g_q^k$ for all $q = 2, \cdots, Q$ and all $k, \ell \in \mathcal{C}$.

These assumptions have interpretations analogous to those given in Section 3.3. Then, Propositions 2 and 3 admit the following generalization.

PROPOSITION 4. *Suppose that assumptions* (**A1E**) *and* (**A2E**) *hold. Then, there exists an optimal solution of problem* (17) *that satisfies* $\mu_j^k = \eta_j^{\ell k} = 0$ *for all* $j > 1$ *and* $k, \ell \in \mathcal{C}$. *Moreover, if* $b_{q,r-q}^k = f_{q,r-q}^{\ell k} = h_q^k = 0$ *for all* $q, r \in \mathcal{Q}, \ell, k \in \mathcal{C}$ *and Assumption (**A3E**) holds, then problem* (17) *reduces to the following linear program:*

$$
\begin{aligned}
& \lambda^* = \min \ \lambda \\
& \text{s.t. } \lambda + \sum_{k \in \mathcal{C}} \Big\{ g^k s_1^k \theta^k + \mathbb{1}_{\{s_1^k > 0\}} a^k \mu^k + \sum_{\ell \in \mathcal{C}} s_{Q+1}^{\ell} \mathbb{1}_{\{s_1^k > 0\}} d^{\ell k} \eta^{\ell k} \Big\} \ge 0 \quad \forall \boldsymbol{s} \in \mathcal{S}^* \\
& \qquad \sum_{k \in \mathcal{C}} \Big\{ c_{\theta^k} \theta^k + c_{\mu^k} \mu^k + \sum_{\ell \in \mathcal{C}} c_{\eta^{\ell k}} \eta^{\ell k} \Big\} \le 1 \\
& \qquad \theta^k, \mu^k, \eta^{\ell k} \ge 0, \ell, k \in \mathcal{C}, \ \lambda \in \mathbb{R},
\end{aligned}
\tag{18}
$$

where $g^k := g_1^k$, $a^k := a_{11}^k$, $d^{\ell k} := d_{11}^{\ell k}$, $\mu^k = \mu_1^k$ *and* $\eta^{\ell k} = \eta_1^{\ell k}$ *for all* $k, \ell \in \mathcal{C}$, *and* $\mathcal{S}^*$ *is defined as*

$$
\mathcal{S}^* := \big\{ \boldsymbol{s} \in \widetilde{\mathcal{S}} \colon \boldsymbol{s}_1^k \in \{0, 1, n^k\}, \boldsymbol{s}_2^k = \boldsymbol{s}_3^k = \cdots = \boldsymbol{s}_Q^k = 0 \ \forall k \in \mathcal{C} \big\}.
$$

Based on this result we conclude that the main properties of optimal solutions of the single-clique model also hold in the multi-clique case. That is, an optimal cascade propagation can be achieved by investing only in transitions of size one ("incremental", or "step-by-step" influence), i.e., it is not necessary to provide resources for multi-step jumps. Furthermore, we can consider only *bottleneck* states $\mathcal{S}^* \subset \mathcal{S}$ (recall our discussion in Section 3.3 next to Proposition 3), that correspond to all the nodes within a clique being either all inactive, or all active except for one inactive.

Observe that the LP (18) has $N(2+N)$ variables and 3^N constraints, which implies orders of magnitude reduction in the problem size with respect to the original program (17). Also, *the size of the problem does not depend on the number of nodes in the network*, but depends only on the number of cliques N. Yet, the number of constraints in (18) is still exponential in N, which makes the problem intractable for networks that consist of a large number of cliques. Nevertheless, it turns out that the special structure of program (18) can be exploited to find an optimal solution in a time that is polynomial in N. The details are discussed in the next section.

4.4. Hierarchical Structure of an Optimal Solution

In this section we discuss a polynomial-time method for finding an optimal solution for the multi-clique model of the preceding section, which reveals a hierarchical tree-like structure of an optimal cascade propagation. Namely, a simple closed-form solution for program (18) can be constructed by means of a *Minimum Spanning Arborescence* (MSA) on an auxiliary graph. An MSA can be thought as the rooted version of a *minimum spanning tree* on directed networks; specifically, T is an MSA rooted at $r \in V$ on network (V, E) if it is a spanning subgraph with $|E|-1$ edges of minimum cost such that there exists a directed path from r to any other node in V, see, e.g., Edmonds [19], Gabow et al. [22]. Note that an MSA can be found in a time that is polynomial in the size of the network (see the above references).

In order to proceed, let for a given $k \in \mathcal{C}$ define the function $F(k)$ as

$$F(k) := \min\Big\{ \frac{c_{\eta^{kk}}}{n^k d^{kk}}, \min_{\ell \in \mathcal{C}, \ell \neq k} \Big\{ \frac{c_{\eta^{\ell k}}}{n^\ell d^{\ell k}} \frac{n^k - 1}{n^k} \Big\} \Big\},$$

and let k' denote an element of $\mathcal{C}$ that attains the above minimum. Also, let

$$m^k := \min\Big\{ \frac{c_{\theta^k}}{g^k}, \frac{c_{\mu^k}}{a^k}, \frac{c_{\theta^k}}{n^k g^k} + F(k) \Big\}, \quad k \in \mathcal{C}. \tag{19}$$

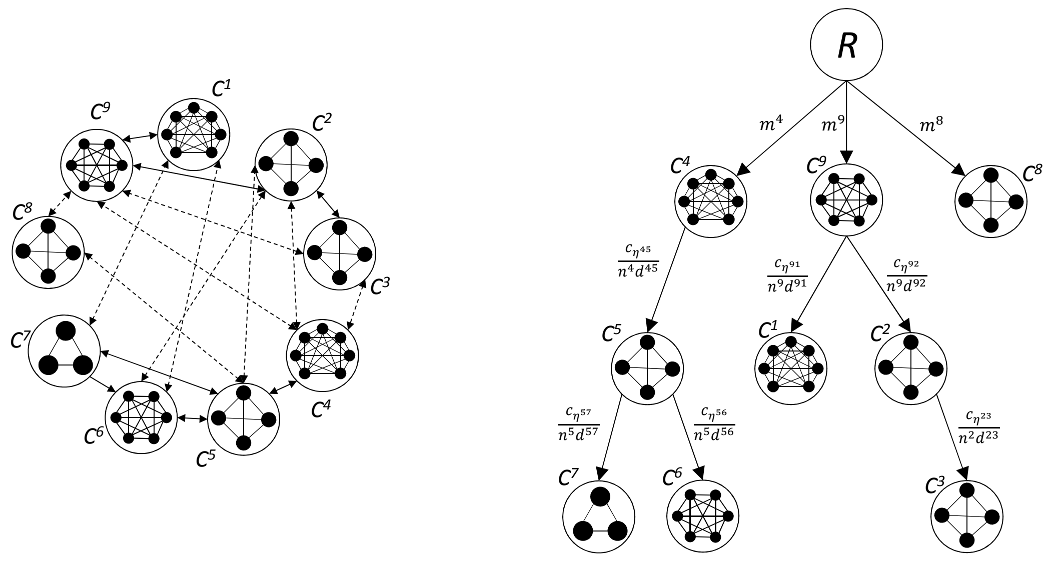
Consider an auxiliary directed graph $\mathcal{G} := (\mathcal{N}, \mathcal{E})$, where the set of nodes is $\mathcal{N} := \mathcal{C} \cup \{R\}$, the set of edges is given by $\mathcal{E} := E^R \cup E^H$, with $E^R := \{(R,k)\colon k \in \mathcal{C}, m^k < \infty\}$ and E^H is the set of edges of the high-level graph, i.e., $E^H := \{(\ell, k)\colon k, \ell \in \mathcal{C}, d^{\ell k} > 0\}$. Observe that the both the auxiliary graph $\mathcal{G}$ and the high-level graph have a similar structure; the difference being that $\mathcal{G}$ has an additional node R, and it also has arcs between R and all the nodes of the high-level graph. The weights (costs) of $\mathcal{G}$ are defined by

$$y_{ij} := \begin{cases} m^k, & \text{if } i = R \text{ and } j = k \\ \frac{c_{\eta^{\ell k}}}{n^\ell d^{\ell k}}, & \text{if } i = \ell,\ j = k. \end{cases}$$

Intuitively, the auxiliary graph $\mathcal{G}$ elucidates the role that each mode of influence has on the cliques in the network. Specifically, observe that the lower is the cost $y_{\ell k}$ of the arc between C^ℓ and C^k, the larger is the influence that C^ℓ has on C^k; hence, the more likely that it is optimal for clique C^ℓ to influence C^k. Similarly, the lower is the cost y_{Rk} of the arc between R and C^k, the more reasonable it seems that it is optimal for the decision-maker to invest into either increasing the external influence or the intrinsic capability rate of C^k (the former if $m^k = c_{\mu^k}/a^k$, and the latter if $m^k < c_{\mu^k}/a^k$). Hence, we might expect that "low cost" connected subgraphs of $\mathcal{G}$ can somehow be mapped into good-quality feasible (or even optimal) solutions of (18). This intuition is formalized in the next Proposition.

Proposition 5. *Assume that* (**A1E**)*,* (**A2E**)*,* (**A3E**) *hold, that* $b^k_{q,r-q} = f^{\ell k}_{q,r-q} = h^k_q = 0$ *for all* $q, r \in \mathcal{Q}, \ell, k \in \mathcal{C}$*, and that there exists a directed path from* R *to all the other nodes in* $\mathcal{G}$*. Let* $\mathcal{T} \subseteq \mathcal{E}$ *be a MSA of* $\mathcal{G}$ *rooted at* R *and let* $c(\mathcal{T})$ *denote its cost, i.e.,* $c(\mathcal{T}) = \sum_{(i,j) \in \mathcal{T}} y_{ij}$*. For any* $k \in \mathcal{C}$*, set*

$$\theta^{k,*} = \begin{cases} \frac{1}{g^k c(\mathcal{T})}, & \textit{if } (R,k) \in \mathcal{T} \textit{ and } m^k = \frac{c_{\theta^k}}{g^k} \\ \frac{1}{n^k g^k c(\mathcal{T})}, & \textit{if } (R,k) \in \mathcal{T} \textit{ and } m^k = \frac{c_{\theta^k}}{n^k g^k} + F(k) \\ 0, & \textit{otherwise,} \end{cases}$$



(a) (b)

Figure 6 Example of a MSA $\mathcal{T}$. In (a) the high-level graph is depicted, each node corresponds to a medium-scale graph. In (b), an MSA structure associated with the auxiliary graph $\mathcal{G}$ of the graph in (a) is depicted. The weights correspond to the weights of $\mathcal{G}$.

$$\mu^{k,*} = \begin{cases} \frac{1}{a^k c(\mathcal{T})}, & \text{if } (R,k) \in \mathcal{T} \text{ and } m^k = \frac{c_{\mu^k}}{a^k} \\ 0, & \text{otherwise.} \end{cases}$$

For any $k, \ell \in \mathcal{C}$ set

$$\eta^{\ell k,*} = \begin{cases} \frac{1}{n^\ell d^{\ell k} c(\mathcal{T})}, & \text{if } (\ell,k) \in \mathcal{T} \\ \frac{F(k)}{c_{\eta^{k'k}} c(\mathcal{T})}, & \text{if } (R,k) \in \mathcal{T},\ m^k = \frac{c_{\theta^k}}{n^k g^k} + F(k), \text{ and } \ell = k' \\ 0, & \text{otherwise,} \end{cases}$$

and let $\lambda^* = -1/c(\mathcal{T})$. Then $\theta^{k,*}$, $\mu^{k,*}$, $\eta^{\ell k,*}$ and λ^* provide an optimal solution of LP (18).

Remark 4. The assumption of the existence of a directed path from R to all the other nodes in $\mathcal{G}$ is made to assure the existence of a spanning arborescence rooted at R. This assumption is made without loss of generality. It can be verified that if there exists $k \in \mathcal{C}$ with no such path, then it is not possible for the nodes in clique C^k to activate and any feasible solution of program (18) must satisfy $\lambda = 0$.

Remark 5. The computation of $\mathcal{T}$ can be done in strongly polynomial time in terms of the number of cliques N. To see why, observe that for each k the computation of m^k requires $O(N \log(N))$ time. On the other hand, given all arc costs, an MSA for $\mathcal{G}$ can be computed in $O(N \log(N) + |\mathcal{E}|)$ time, see Gabow et al. [22]. Therefore, $\mathcal{T}$ can be computed in $O(N^2 \log(N))$ time.

Proposition 5 provides a number of important insights into the structure of the optimal solution (see an illustrative example in Figure 6):

<u>*MSA hierarchy and optimal resource allocation:*</u> Observe that the cliques can be classified according to their *level* v in the MSA structure, with $v=1$ and $v=M$ corresponding to the top and bottom levels, respectively. That is, $\mathcal{C}$ can be partitioned as

$$\mathcal{C} = \bigcup_{v=1}^{M} L_v,$$

where L_v is the set of cliques in level v of the hierarchy of $\mathcal{T}$ such that $L_v := \{k \in \mathcal{C} \colon$ there are v arcs in the path from R to k in $\mathcal{T}\}$. In general, the position of a clique in this hierarchy depends on three factors: (i) how reasonable it is to invest into the intrinsic capability of the clique or provide it with the external influence; (ii) how much influence does the clique have on other cliques in the graph, i.e., how "central" is the clique, which is determined by the high-level graph weights and the cost-adjusted network influence rates; and (iii) the clique size n^k.

Cliques in L_1 are the only ones that receive *"independent help."* We introduce this notion to simplify our further discussion. Namely, in this context, independent help means that the resources are allocated to either the clique's intrinsic capability rate ($\theta^k > 0$), or the clique's external influence rate ($\mu^k > 0$). Which one of the two is non-zero depends on the value attained by the minimum defined by m^k. If m^k is given by c_{μ^k}/a^k (thus, external influence, in a sense, is "cheaper" than intrinsic capability), then it is optimal to invest into the external influence rate of C^k; in any other case, it is optimal to invest into its intrinsic effort rate. Furthermore, if $m^k = \frac{c_{\theta^k}}{n^k g^k} + F(k)$, then C^k receives investments into its intrinsic effort along with the network influence from the nodes in $C^{k'}$. Also, it is proven (see the proof of Proposition 5) that if this is the case, then $C^{k'}$ does not receive independent help but receives network influence from C^k.

Cliques beyond the first level (i.e., cliques in L_v with $v \geq 2$) receive investments only for the network influence from the active nodes in other cliques. Specifically, they only receive network influence from nodes in its (unique) parent clique in L_{v-1}. That is, for C^k it is optimal to invest only into $\eta^{\ell(k)k}$, where $\ell(k)$ denotes the unique element of $\mathcal{C}$ such that $(\ell(k), k) \in \mathcal{T}$.

<u>*Independent influence processes in the network:*</u> The hierarchical structure of $\mathcal{T}$ has a number of additional implications. First, it means that the cliques of the network can be naturally divided into *independent clusters*. Indeed, for any $k \in L_1$, define $\mathcal{V}^k$ as the set/cluster of cliques that are reachable from k in $\mathcal{T}$, i.e.,

$$\mathcal{V}^k := \{\ell \in \mathcal{C} \colon \text{ there is a directed path from } k \text{ to } \ell \text{ in } \mathcal{T}\} \cup \{k\}.$$

Observe that $\{V_k\}_{\{k:\,(R,k)\in\mathcal{T}\}}$ is a partition of $\mathcal{C}$ and that cliques from different clusters of the partition do not influence each other; hence, their influence processes are independent. Second, observe that the higher the clique is in the hierarchy, the more likely it is that its nodes activate early on; in particular, nodes in a clique in level $v \geq 2$ can become active only if there is at least an active node in its predecessor clique. Third, the cliques in the first level of $\mathcal{T}$ can be viewed as the *seeds* of the cascade process as they are the only cliques receiving the independent help, and they propagate influence through their active nodes along the other cliques in their cluster. In this sense, these cliques can be thought as "optimal" propagation seeds if the influence problem is framed in the more traditional IM perspective of Kempe et al. [30].

Using the definition of the costs of $\mathcal{G}$ and Proposition 5, it is possible to characterize the cliques for which it is optimal to provide independent help. Intuitively, a clique is in L_1 if it has a large number of nodes, or if its cost-adjusted intrinsic capability (g^k/c_{θ^k}) or external help (a^k/c_{μ^k}) rates are sufficiently large when compared to the adjusted network influence ($n^\ell d^{\ell k}/c_{\eta^{\ell k}}$) rates that other cliques can yield. This intuition is formalized as follows:

Corollary 2. *Suppose the assumptions of Proposition 5 hold. If $k \in \mathcal{C}$ satisfies*

$$\frac{n^\ell d^{\ell k}}{c_{\eta^{\ell k}}} \leq \frac{a^k}{c_{\mu^k}} \quad \forall \ell \in \mathcal{C}\setminus\{k\}, \ \text{or}\ \frac{n^\ell d^{\ell k}}{c_{\eta^{\ell k}}} \leq \frac{g^k}{c_{\theta^k}} \quad \forall \ell \in \mathcal{C}\setminus\{k\},$$

or

$$\frac{n^\ell d^{\ell k}(c_{\theta^k} n^{k'} d^{k'k} - c_{\eta^{k'k}} g^k)}{g^k(c_{\eta^{\ell k}} n^{k'} d^{k'k} - c_{\eta^{k'k}} n^\ell d^{\ell k})} \leq n^k \qquad \forall \ell \in \mathcal{C}\setminus\{k\},\ \text{if}\ k' \neq k$$

or

$$\frac{n^\ell d^{\ell k}(c_{\theta^k} d^{kk} + c_{\eta^{kk}} g^k)}{g^k c_{\eta^{\ell k}} d^{kk}} \leq n^k \qquad \forall \ell \in \mathcal{C}\setminus\{k\},\ \text{if}\ k' = k,$$

then $k \in L_1$, i.e., it is optimal to provide independent help for clique C^k.

In a similar way, it is possible to derive sufficient conditions for a clique C^k to not be in L_1 in an optimal solution, see Corollary 4 in the Appendix. In addition, classical sufficient conditions for a spanning arborescence to be an MSA might allow for determining the optimal solution of (18) without the need for additional computations. We repeat this condition in terms of network $\mathcal{G}$ as follows, see the proof in Karp [29]:

Proposition 6. *For any $k \in \mathcal{C}$ let $k^* \in \mathcal{N}$ be such that $y_{k^*k} \leq y_{\ell k}$ for all $(\ell,k) \in \mathcal{E}$ and define $\mathcal{T}^* = \{(\ell,k) \colon \ell = k^*\ \forall k \in \mathcal{C}\}$. If $\mathcal{T}^*$ is a spanning arborescence of $\mathcal{G}$ rooted at R, then it is an MSA over $\mathcal{G}$ rooted at R.*

This proposition implies that if for every node in $\mathcal{N}$ (with the exception of R) we consider the incoming arc with lowest cost and then form a subnetwork consisting of those arcs, this subnetwork is an MSA as long as it is a spanning arborescence rooted at R. Using this result, we can easily obtain optimal solutions for particular classes of networks. As an example, we have the following simple optimal solution of (18) when $\mathcal{G}\setminus\{R\}$ is strongly connected and all adjusted costs are the same.

COROLLARY 3. *Assume, without loss of generality, that that $n^1 > n^2 \geq \ldots \geq n^N$. Also, suppose that there exist $a,g,d>0$ such that*

$$\frac{a^k}{c_{\mu^k}} = a, \qquad \frac{g^k}{c_{\theta^k}} = g \qquad \forall k\in\mathcal{C}, \qquad \frac{d^{\ell k}}{c_{\eta^{\ell k}}} = d, \qquad \forall k,\ell\in\mathcal{C},$$

and let $\hat{k} = \min\{k\in\mathcal{C}\setminus\{1\} : 1/(dn^1) < m^k\}$. Then

$$\hat{\mathcal{T}} = \bigcup_{k=1}^{\hat{k}-1}\{(R,k)\} \cup \bigcup_{k=\hat{k}}^{N}\{(1,k)\}$$

is an MSA rooted at R of $\mathcal{G}$.

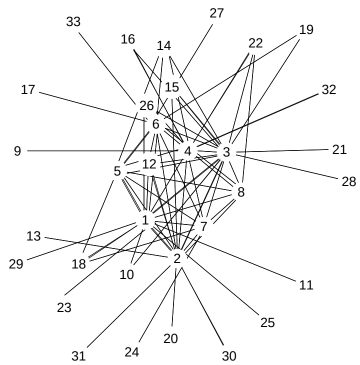
This corollary implies that if all adjusted costs are the same and the network is strongly connected, then the clique size becomes the main decisive factor, i.e., an optimal solution of (18) is to provide independent help to the larger cliques (i.e., cliques $1,2,\ldots,\hat{k}-1$), and then the largest clique should influence the smaller cliques. This result, thus, confirms the intuition regarding the role of large clusters of nodes (in connected networks) in terms of influence: they should be the ones at which independent help is directed, in order for them to spread the influence further across the rest of the network.

5. A Numerical Illustration

In this section we present an example that illustrates the general model of Section 4 and its defining components. In particular, we show how the sizes of the cliques, the values of the model coefficients, and the location (or "centrality") of the cliques play a role in optimal solutions.

Network generation. We generate network G with $n = 100$ nodes using the "small-world" BA structure (Albert and Barabási [2]). The initial network in the generation process was a uniform random graph (Erdős and Rényi [21]) with $n_0 = 75$ nodes and $p_0 = 3/4$. We partition the network into cliques by successively identifying and removing the largest clique. The high-level network weights are computed according to (14). The resulting high-level graph has $N = 33$ cliques, see Figure 7(a). The size of the cliques and the corresponding weights for the largest twelve cliques are given in Figure 7(b).

Data instances. We consider four different settings for the values of a^k, g^k and $d^{\ell k}$. In *Instance A* the values of these parameters are fixed at 1. *Instances B*, *C* and *D* are obtained from *Instance A* by setting $a^k = 2$, $g^k = 2$ and $d^{\ell k} = \ell$, respectively, for all $k\in\mathcal{C}$.



(a) High-level graph overview; each node is a clique by itself.

$C^\ell \| C^k$	1	2	3	4	5	6	7	8
1 (16)	1	0.75	0.73	0.71	0.73	0.64	0.72	0.79
2 (14)		1	0.75	0.72	0.69	0.71	0.75	0.64
3 (12)			1	0.77	0.67	0.71	0.65	0.64
4 (10)				1	0.69	0.69	0.70	0.77
5 (9)					1	0.67	0.67	0.78
6 (7)						1	0.68	0.76
7 (4)							1	0.58
8 (3)								1

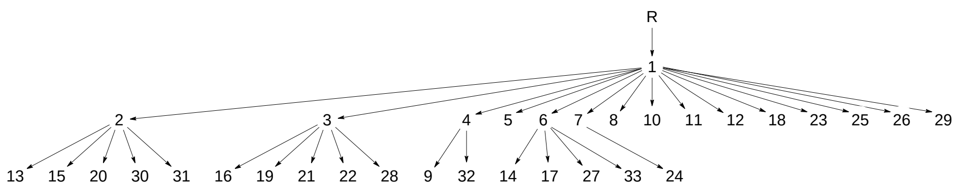
(b) Weights between the largest eight cliques out of $N = 33$. The sizes are given by the number in parenthesis in the first column.

Figure 7 The high-level network used in the computational experiments.

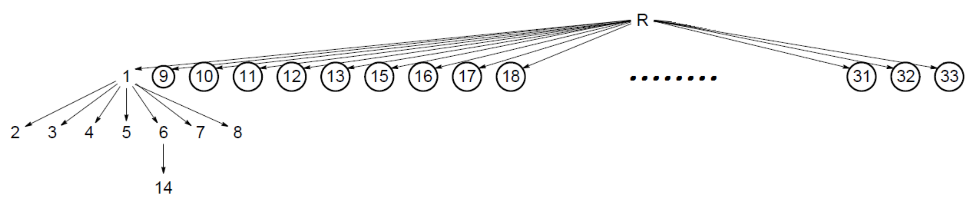
Results and analysis. Figure 8 displays the MSA structures $\mathcal{T}$ for each of the four instances. In particular, *Instance A* (see Figure 8(a)) demonstrates the importance of the clique sizes. Observe that the largest clique (C^1) is the only one receiving independent help, and that it further distributes influence to the other cliques. Particularly, the largest cliques (C^2, …, C^6) are in the second level of the hierarchy, while the smaller ones are at the bottom. This behavior changes when the effectiveness of external influence (or intrinsic capabilities) is doubled in *Instance B* (respectively, *Instance C*), see Figure 8(b). Here, smaller cliques are no longer at the bottom but at the top, as any investment in the external influence (or intrinsic capabilities) is twice as effective with respect to *Instance A*. Observe, however, that most of the structure present in the solution of *Instance A* (Figure 8(a)) for the largest cliques repeats itself in *Instances B* and *C* (Figure 8(b)), suggesting that these cliques are less sensitive to changes of the independent help rates, i.e., the connections between each other are stronger.

In *Instance D* (Figure 8(c)) it can be seen that while the clique size is important, it is not the only key factor behind the spread of influence. In this case, the smaller the clique size the more influence it has on other cliques (recall $d^{\ell k} = \ell$). Thus, e.g., influence from C^{30} is thirty times more effective than influence from C^1. In this setting, cliques of intermediate size (C^3, C^4 and C^5) are the first in the hierarchy that distributes the influence to the rest of the network, while C^1 and C^2 are almost at the bottom. Also, it is surprising that cliques with a very high influence potential (i.e., C^{21}, …, C^{33}) do not influence any clique in the network. Intuitively, this follows from their lack of centrality or "connectedness" in the overall network structure, see Figure 7(a), and illustrates that only a very high influence is not sufficient for actors in a network to spread influence; they need to be closely connected to the rest of the network.

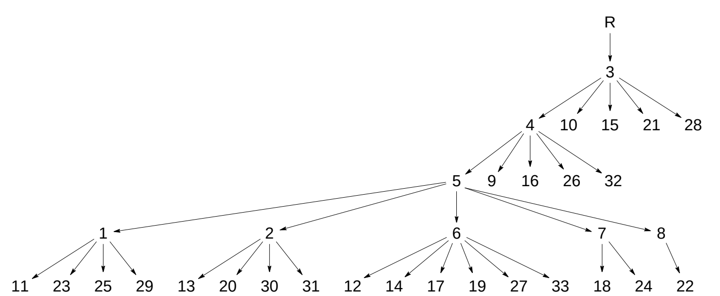
We close by noting that in this example, increasing the external influence (*Instance B*) has similar effects as increasing the intrinsic capability parameters (*Instance C*). This is to be expected, see equation (19). Also, with the exception of *Instances B* and *C*, the independent help that the cliques in the first level receive is the intrinsic capability plus network influence from the active nodes, i.e., in $m^k = \frac{c_{\theta k}}{n^k g^k} + F(k)$. As these cliques are relatively large, this observation again indicates the importance of the network effect, i.e., receiving network influence from the active nodes results in faster activation times that scale with the size of the network (recall the discussions in Sections 3.3 and 3.4).



(a) MSA for Instance A



(b) MSA for Instances B & C



(c) MSA for Instance D

Figure 8 MSA structures for the optimal solutions. Cliques in the first level within circles indicate that the independent help is either all external influence (Instance B), or all intrinsic capability (Instance C).

6. Conclusions

In this work we developed a stochastic framework to minimize the time at which cascading processes propagate through a network. Within the proposed approach, the cascading processes are modeled using multiscale graphs, where the time dynamics of the cascade are described by a Markov chain. The evolution of this process depends on the multiscale structure of the network, the costs associated with the parameters governing the Markov chain, and the state of the cascade at each node of the network.

By considering that there are inherent cost associated with changing the parameters of the Markov chain, the optimization is formulated as a resource allocation problem with the objective to determine the parametrization of the process that minimizes the time until the cascade reaches all the nodes in the network. We studied in detail a particular model that assumes that the transition rates are linear functions of the control parameters, and showed how the corresponding optimization problem reduces to a linear program.

Under natural assumptions on the relationship between the parameters and costs of the model, we derived closed-form analytical expressions for the optimal resource allocation. This model elucidates several properties of cascade propagation in networks. Namely, it shows the relevance of the network effect for the rate at which the cascade spreads, reveals that an "incremental" propagation can be as optimal as "big-leaps" spread, and shows that *bottleneck* states representing that all nodes in the cliques are either all inactive, or active (except one), are responsible for modulating the rate at which the cascade propagates.

Moreover, we established the relationship that exists between the cascade propagation optimization problem and a minimum spanning arborescence (a combinatorial optimization problem) on an auxiliary graph. From this connection, an optimal solution can be computed in a time that is polynomial in the size of the network. In addition, the obtained solution yields a hierarchical classification of the network clusters. This classification is valuable for identifying clusters that can be used as the seeds of the cascade process, the determining the relative relevance of each clusters for the cascade propagation, and provides a natural division of the influence process into independent classes of cliques. In particular, our model showed that large cliques should be seeds of the cascade process in settings where the parameters defining the influence spread are homogenous across the network.

Acknowledgments

This material is based upon research supported by the the grant from the Defense Threat Reduction Agency (DTRA) HDTRA1-14-1-0065 and the U.S. Air Force Research Laboratory (AFRL) Mathematical Modeling and Optimization Institute. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon.

References

[1] Ackerman E, Ben-Zwi O, Wolfovitz G (2010) Combinatorial model and bounds for target set selection. *Theoretical Computer Science* 411(44):4017–4022.

[2] Albert R, Barabási AL (2002) Statistical mechanics of complex networks. *Reviews of modern physics* 74(1):47.

[3] Amin M (2000) National infrastructures as complex interactive networks. *Automation, Control, and Complexity: An Integrated Approach* 263–286.

[4] Aral S, Walker D (2014) Tie strength, embeddedness, and social influence: A large-scale networked experiment. *Management Science* 60(6):1352–1370.

[5] Asavathiratham C, Roy S, Lesieutre B, Verghese G (2001) The influence model. *Control Systems, IEEE* 21(6):52–64.

[6] Balasundaram B, Butenko S, Hicks IV (2011) Clique relaxations in social network analysis: The maximum k-plex problem. *Operations Research* 59(1):133–142.

[7] Bapna R, Umyarov A (2015) Do your online friends make you pay? a randomized field experiment on peer influence in online social networks. *Management Science* 61(8):1902–1920.

[8] Ben-Zwi O, Hermelin D, Lokshtanov D, Newman I (2011) Treewidth governs the complexity of target set selection. *Discrete Optimization* 8(1):87–96.

[9] Bhargava S, Kumar A, Mukherjee A (1993) A stochastic cellular automata model of innovation diffusion. *Technological forecasting and social change* 44(1):87–97.

[10] Blume L, Easley D, Kleinberg J, Kleinberg R, Tardos É (2011) Which networks are least susceptible to cascading failures? *Foundations of Computer Science (FOCS), 2011 IEEE 52nd Annual Symposium on*, 393–402 (IEEE).

[11] Brummitt CD, DSouza RM, Leicht E (2012) Suppressing cascades of load in interdependent networks. *Proceedings of the National Academy of Sciences* 109(12):E680–E689.

[12] Buttle FA (1998) Word of mouth: understanding and managing referral marketing. *Journal of strategic marketing* 6(3):241–254.

[13] Chen W, Lakshmanan LV, Castillo C (2013) Information and influence propagation in social networks. *Synthesis Lectures on Data Management* 5(4):1–177.

[14] Chen W, Lu W, Zhang N (2012) Time-critical influence maximization in social networks with time-delayed diffusion process. *arXiv preprint arXiv:1204.3074* .

[15] Chen W, Wang C, Wang Y (2010) Scalable influence maximization for prevalent viral marketing in large-scale social networks. *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, 1029–1038 (ACM).

[16] Domingos P, Richardson M (2001) Mining the network value of customers. *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, 57–66 (ACM).

[17] Dong W, Pentland A (2006) Multi-sensor data fusion using the influence model. *Wearable and Implantable Body Sensor Networks, 2006. BSN 2006. International Workshop on*, 4–pp (IEEE).

[18] Duenas-Osorio L, Vemuru SM (2009) Cascading failures in complex infrastructure systems. *Structural safety* 31(2):157–167.

[19] Edmonds J (1967) Optimum branchings. *Journal of Research of the National Bureau of Standards B* 71(4):233–240.

[20] Eftekhar M, Ganjali Y, Koudas N (2013) Information cascade at group scale. *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, 401–409 (ACM).

[21] Erdős P, Rényi A (1959) On random graphs. *Publicationes Mathematicae Debrecen* 6:290–297.

[22] Gabow HN, Galil Z, Spencer T, Tarjan RE (1986) Efficient algorithms for finding minimum spanning trees in undirected and directed graphs. *Combinatorica* 6(2):109–122.

[23] Goel S, Anderson A, Hofman J, Watts DJ (2016) The structural virality of online diffusion. *Management Science* 62(1):180–196.

[24] Gomez Rodriguez M, Balduzzi D, Schölkopf B, Scheffer GT, et al. (2011) Uncovering the temporal dynamics of diffusion networks. *28th International Conference on Machine Learning (ICML 2011)*, 561–568 (International Machine Learning Society).

[25] Goyal A, Bonchi F, Lakshmanan LV, Venkatasubramanian S (2013) On minimizing budget and time in influence propagation over social networks. *Social Network Analysis and Mining* 3(2):179–192.

[26] Hill S, Provost F, Volinsky C (2006) Network-based marketing: Identifying likely adopters via consumer networks. *Statistical Science* 256–276.

[27] Jeffcoat DE, Krokhmal PA, Zhupanska OI (2006) Effects of cueing in cooperative search. *Naval Research Logistics (NRL)* 53(8):814–821.

[28] Jun M, Jeffcoat DE (2005) Convergence properties of continuous-time markov chains with application to target search. Technical report, DTIC Document.

[29] Karp RM (1971) A simple derivation of edmonds' algorithm for optimum branchings. *Networks* 1(3):265–272.

[30] Kempe D, Kleinberg J, Tardos É (2003) Maximizing the spread of influence through a social network. *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, 137–146 (ACM).

[31] Kempe D, Kleinberg J, Tardos É (2015) Maximizing the spread of influence through a social network. *Theory of Computing* 11(4):105–147.

[32] Kim DH, Kim BJ, Jeong H (2005) Universality class of the fiber bundle model on complex networks. *Physical review letters* 94(2):025501.

[33] Kulkarni VG (2009) *Modeling and analysis of stochastic systems* (Boca Raton,FLf: CRC Press).

[34] Lewis AS, Overton ML (1996) Eigenvalue optimization. *Acta numerica* 5:149–190.

[35] Little RG (2002) Controlling cascading failure: understanding the vulnerabilities of interconnected infrastructures. *Journal of Urban Technology* 9(1):109–123.

[36] Liu B, Cong G, Xu D, Zeng Y (2012) Time constrained influence maximization in social networks. *ICDM*, 439–448.

[37] Ma H, Li H, Song JB (2013) Influence models of cascading failure and frequency oscillation in the power grid. *Innovative Smart Grid Technologies (ISGT), 2013 IEEE PES*, 1–6 (IEEE).

[38] Majdandzic A, Podobnik B, Buldyrev SV, Kenett DY, Havlin S, Stanley HE (2014) Spontaneous recovery in dynamical networks. *Nature Physics* 10(1):34–38.

[39] Peng S, Li Y, Zheng B (2008) Steady states and critical behavior of epidemic spreading on complex networks. *Intelligent Control and Automation, 2008. WCICA 2008. 7th World Congress on*, 3481–3486 (IEEE).

[40] Plateau B, Atif K (1991) Stochastic automata network of modeling parallel systems. *Software Engineering, IEEE Transactions on* 17(10):1093–1108.

[41] Richardson M, Domingos P (2002) Mining knowledge-sharing sites for viral marketing. *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, 61–70 (ACM).

[42] Richoux WJ, Verghese GC (2011) A generalized influence model for networked stochastic automata. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on* 41(1):10–23.

[43] Roy S, Lesieutre BC, Verghese GC (2002) Resource allocation in networks: a case study of the influence model. *System Sciences, 2002. HICSS. Proceedings of the 35th Annual Hawaii International Conference on*, 875–884 (IEEE).

[44] Samadi M (2016) *Optimal Strategies for Controlling Cascades in Social Networks: An In uence Maximization Approach*. Ph.D. thesis, STATE UNIVERSITY OF NEW YORK AT BUFFALO.

[45] Shaked M, Shanthikumar JG (1994) *Stochastic orders and their applications* (Academic Pr).

[46] Sun J, Boyd S, Xiao L, Diaconis P (2006) The fastest mixing markov process on a graph and a connection to a maximum variance unfolding problem. *SIAM review* 48(4):681–699.

[47] Wang Y, Cong G, Song G, Xie K (2010) Community-based greedy algorithm for mining top-k influential nodes in mobile social networks. *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, 1039–1048 (ACM).

[48] Wasserman S, Faust K (1994) *Social network analysis: Methods and applications*, volume 8 (Cambridge university press).

[49] Watts DJ (2002) A simple model of global cascades on random networks. *Proceedings of the National Academy of Sciences* 99(9):5766–5771.

[50] Watts DJ, Dodds PS (2007) Influentials, networks, and public opinion formation. *Journal of consumer research* 34(4):441–458.

[51] West DB, et al. (2001) *Introduction to graph theory*, volume 2 (Prentice hall Upper Saddle River).

[52] Wu ZX, Peng G, Wang WX, Chan S, Wong EWM (2008) Cascading failure spreading on weighted heterogeneous networks. *Journal of Statistical Mechanics: Theory and Experiment* 2008(05):P05013.

Appendix. Proof of Main Results.

We provide the proofs of Propositions 1, 4 and 5 for the general model of Section 4. We omit the proofs of Propositions 2 and 3 (which deal with the complete network case of Section 3) as these propositions are a direct consequence of the results for general networks. The proofs of the corollaries are at the end of this section.

Proof of Proposition 1. In the state space $\mathcal{S}$ of $Y(\boldsymbol{\zeta})$ impose a *reverse* lexicographical order, that is, $(\boldsymbol{s}^1, \boldsymbol{s}^2, \cdots, \boldsymbol{s}^N) < (\boldsymbol{i}^1, \boldsymbol{i}^2, \cdots, \boldsymbol{i}^N)$ if and only if there exist $\ell, r > 0$ such that $\boldsymbol{s}^k = \boldsymbol{i}^k$ for all $k < \ell$, $s^\ell_q = i^\ell_q$ for all $q < r$ and $s^\ell_r > i^\ell_r$. It is readily verified that this order is a total order over $\mathcal{S}$, and that if in the generator $\boldsymbol{\Xi}(\boldsymbol{\zeta})$ the states are ordered in this fashion, then $\boldsymbol{\Xi}(\boldsymbol{\zeta})$ is a upper triangular matrix.

It can be concluded that the eigenvalues of $\boldsymbol{\Xi}(\boldsymbol{\zeta})$ are given by its diagonal elements and as such all of them are negative except for one that is zero (associated with the row of state $\boldsymbol{s}_A$). Now, let $J_1(\boldsymbol{\zeta})$, $J_2(\boldsymbol{\zeta})$, …, $J_m(\boldsymbol{\zeta})$ be the Jordan blocks of the Jordan normal form of $\boldsymbol{\Xi}(\boldsymbol{\zeta})$, let $\lambda_1(\boldsymbol{\zeta})$, …, $\lambda_m(\boldsymbol{\zeta})$ be the corresponding eigenvalue of each block, and let $j_1, \ldots, j_m$ be the size of each block, i.e., the algebraic multiplicity of each eigenvalue. Without loss of generality, assume that they are organized in decreasing order, so $0 = \lambda_1(\boldsymbol{\zeta}) > \lambda_2(\boldsymbol{\zeta}) > \ldots > \lambda_m(\boldsymbol{\zeta})$ and note that these eigenvalues are precisely the eigenvalues of $\boldsymbol{\Xi}(\boldsymbol{\zeta})$ as $\boldsymbol{\Xi}(\boldsymbol{\zeta})$ and is similar to its Jordan canonical form. The same holds true for the multiplicity of the eigenvalues. By virtue of Theorem 1 and Proposition 3 of [28], it can be concluded that there exist non-negative $\tilde{M}_{ki}(\boldsymbol{\zeta})$ such that

$$P[T(\boldsymbol{\zeta}) \le t] \ge 1 - \sum_{i=2}^{m} e^{\lambda_i(\boldsymbol{\zeta})t} \sum_{k=1}^{j_i} \frac{t^{k-1}}{(k-1)!} \tilde{M}_{ki}(\boldsymbol{\zeta})$$

and hence, since $\lambda_k(\boldsymbol{\zeta}) < \lambda_2(\boldsymbol{\zeta}) < 0$ for $k = 3, \ldots, m$, there exist a constant $M_{12}(\boldsymbol{\zeta}) > 0$ and a time $t_0(\boldsymbol{\zeta}) > 0$ such that for all $t \ge t_0(\boldsymbol{\zeta})$

$$e^{\lambda_2(\boldsymbol{\zeta})t} \frac{t^{j_2-1}}{(j_2-1)!} M_{12}(\boldsymbol{\zeta}) \ge \sum_{i=3}^{m} e^{\lambda_i(\boldsymbol{\zeta})t} \sum_{k=1}^{j_i} \frac{t^{k-1}}{(k-1)!} \tilde{M}_{ki}(\boldsymbol{\zeta}).$$

It can be concluded that there exist a constant $M(\boldsymbol{\zeta}) > 0$ such that for all $t \ge t_0(\boldsymbol{\zeta})$

$$\begin{aligned} P[T(\boldsymbol{\zeta}) \le t] &\ge 1 - e^{\lambda_2(\boldsymbol{\zeta})t} t^{j_2-1} M(\boldsymbol{\zeta}) \\ &= 1 - e^{\lambda_2(\boldsymbol{\zeta})t} t^{K(\boldsymbol{\zeta})-1} M(\boldsymbol{\zeta}) \end{aligned} \tag{20}$$

where (20) follows from the definition of $K(\boldsymbol{\zeta})$.

Proof of Proposition 4. Observe the dual of (17) is given by

$$\max -\delta - \sum_{\boldsymbol{s} \in \tilde{\mathcal{S}}} K_{\boldsymbol{s}} \rho_{\boldsymbol{s}} \tag{21}$$

$$
\begin{aligned}
\text{s.t.}\ & -\delta+\sum_{\boldsymbol{s}\in\tilde{\mathcal{S}}}\frac{A^k_{\boldsymbol{s}j}}{c_{\mu^k_j}}\rho_{\boldsymbol{s}}\le 0 && \forall j\in\mathcal{J},k\in\mathcal{C}\\
& -\delta+\sum_{\boldsymbol{s}\in\tilde{\mathcal{S}}}\frac{s^\ell_{Q+1}D^k_{\boldsymbol{s}j}}{c_{\eta^{\ell k}_j}}\rho_{\boldsymbol{s}}\le 0 && \forall j\in\mathcal{J},k,\ell\in\mathcal{C}\\
& -\delta+\sum_{\boldsymbol{s}\in\tilde{\mathcal{S}}}\frac{G^k_{\boldsymbol{s}}}{c^k_\theta}\rho_{\boldsymbol{s}}\le 0 && \forall k\in\mathcal{C}\\
& \sum_{\boldsymbol{s}\in\tilde{\mathcal{S}}}\rho_{\boldsymbol{s}}=1\\
& \rho_{\boldsymbol{s}}\ge 0,\ \forall \boldsymbol{s}\in\tilde{\mathcal{S}},\delta\ge 0.
\end{aligned}
$$

We claim that this program is equivalent to

$$
\begin{aligned}
\max\ & -\delta-\sum_{\boldsymbol{s}\in\tilde{\mathcal{S}}}K_{\boldsymbol{s}}\rho_{\boldsymbol{s}} \\
\text{s.t.}\ & -\delta+\sum_{\boldsymbol{s}\in\tilde{\mathcal{S}}}\frac{A^k_{\boldsymbol{s}}}{c^k_{\mu}}\rho_{\boldsymbol{s}}\le 0 && \forall k\in\mathcal{C}\\
& -\delta+\sum_{\boldsymbol{s}\in\tilde{\mathcal{S}}}\frac{s^\ell_{Q+1}D^k_{\boldsymbol{s}}}{c^{\ell k}_{\eta}}\rho_{\boldsymbol{s}}\le 0 && \forall k,\ell\in\mathcal{C}\\
& -\delta+\sum_{\boldsymbol{s}\in\tilde{\mathcal{S}}}\frac{G^k_{\boldsymbol{s}}}{c^k_\theta}\rho_{\boldsymbol{s}}\le 0 && \forall k\in\mathcal{C}\\
& \sum_{\boldsymbol{s}\in\tilde{\mathcal{S}}}\rho_{\boldsymbol{s}}=1\\
& \rho_{\boldsymbol{s}}\ge 0,\ \forall \boldsymbol{s}\in\tilde{\mathcal{S}},\delta\ge 0.
\end{aligned}
\tag{22}
$$

where for notational simplicity $A^k_{\boldsymbol{s}}:=A^k_{\boldsymbol{s}1}$, $D^k_{\boldsymbol{s}}:=D^k_{\boldsymbol{s}1}$, $c^k_\mu:=c^k_{\mu_1}$ and $c^{\ell k}_\eta:=c^{\ell k}_{\eta_1}$. Indeed, first observe that any point feasible in (21) is also feasible in (22). To prove the converse, pick $j\in\mathcal{J}$, $k\in\mathcal{C}$ and let $\boldsymbol{s}\in\tilde{\mathcal{S}}$ be arbitrary. Note that

$$
A^k_{\boldsymbol{s}j}=\sum_{q=1}^{Q-j+1}\mathbb{1}_{\{s^k_q>0\}}a^k_{qj}\ge\sum_{q=1}^{Q-j}\mathbb{1}_{\{s^k_q>0\}}a^k_{q,j+1}=A^k_{\boldsymbol{s},j+1},
$$

where the inequality follows as the summand in the left has one more term and since $a^k_{qj}\ge a^k_{q,j+1}$ from assumption (**A2E**). Moreover, since by assumption (**A1E**), $c_{\mu_j}\le c_{\mu_{j+1}}$, we can conclude that

$$
\frac{A^k_{\boldsymbol{s},j}}{c_{\mu^k_j}}\ge\frac{A^k_{\boldsymbol{s},j+1}}{c_{\mu^k_{j+1}}}.
$$

As this holds for any arbitrary $\boldsymbol{s}\in\tilde{\mathcal{S}}$ and $k\in\mathcal{C}$ it follows that $\sum_{\boldsymbol{s}\in\tilde{\mathcal{S}}}\frac{A^k_{\boldsymbol{s}j}}{c_{\mu^k_j}}\rho_{\boldsymbol{s}}\ge\sum_{\boldsymbol{s}\in\tilde{\mathcal{S}}}\frac{A^k_{\boldsymbol{s}j+1}}{c^k_{\mu_{j+1}}}\rho_{\boldsymbol{s}}$ for all $k\in\mathcal{C}$. Using the same arguments, it is shown that $\sum_{\boldsymbol{s}\in\tilde{\mathcal{S}}}\frac{s^\ell_{Q+1}D^k_{\boldsymbol{s}j}}{c_{\eta^{\ell k}_j}}\rho_{\boldsymbol{s}}\ge\sum_{\boldsymbol{s}\in\tilde{\mathcal{S}}}\frac{s^\ell_{Q+1}D^k_{\boldsymbol{s}j+1}}{c_{\eta^{\ell k}_{j+1}}}\rho_{\boldsymbol{s}}$ for all

$k,\ell \in \mathcal{C}$. Therefore, it can be concluded that any point (δ,ρ) feasible in (22) it is also feasible in (21), which yields the desired converse.

As both programs (21) and (22) have the same feasible region and the same objective function, then they are equivalent. Now, consider the dual of (22):

$$\begin{aligned} &\min \lambda \\ &\text{s.t. } \lambda + \sum_{k\in\mathcal{C}} \Big\{ G_{\boldsymbol{s}}^k \theta^k + A_{\boldsymbol{s}}^k \mu^k + \sum_{\ell\in\mathcal{C}} s_{Q+1}^{\ell} D_{\boldsymbol{s}}^{\ell k} \eta^{\ell k} \Big\} \geq -K_{\boldsymbol{s}} \quad \forall \boldsymbol{s} \in \widetilde{\mathcal{S}} \\ &\quad -\sum_{k\in\mathcal{C}} \Big\{ c_{\theta^k}\theta^k + c_{\mu^k}\mu^k + \sum_{\ell\in\mathcal{C}} c_{\eta^{\ell k}} \eta^{\ell k} \Big\} \geq -1, \\ &\quad \mu^k, \eta^{\ell k}, \theta^k \geq 0 \ \forall k,\ell \in \mathcal{C}, \lambda \in \mathbb{R}. \end{aligned} \tag{23}$$

Program (21) is equivalent to (22) and program (23) is the dual of (22). It follows that program (23) is equivalent to (17). Since in program (23) all the variables μ_j^k, η_j^k $j \in \mathcal{J}$ and $j > 1$ do not exist, for all $k \in \mathcal{C}$, it can be concluded that they are zero.

We now proceed to prove that problem (23) reduces to problem (18). To this end, first observe that from the assumptions of linearity (i.e., that $h_q^k = b_q^k = f_q^{\ell k} = 0$), then $K_{\boldsymbol{s}} = 0$ for all $\boldsymbol{s} \in \widetilde{\mathcal{S}}$. Now, for any $\boldsymbol{s} \in \tilde{S}$ and any λ, $\boldsymbol{\zeta}$ define

$$c(\boldsymbol{s},\lambda,\boldsymbol{\zeta}) := \lambda + \sum_{k\in\mathcal{C}} \Big\{ G_{\boldsymbol{s}}^k \theta^k + A_{\boldsymbol{s}}^k \mu^k + \sum_{\ell\in\mathcal{C}} s_{Q+1}^{\ell} D_{\boldsymbol{s}}^{\ell k} \eta^{\ell k} \Big\},$$

and let

$$\bar{\mathcal{S}} = \{\boldsymbol{s} \in \mathcal{S} \colon s_1^k \in \{0,1,\dots,n^k\}, s_2^k = s_3^k = \ldots = s_Q^k = 0 \ \forall k \in \mathcal{C}\}.$$

Define

$$F_1 = \{(\lambda,\boldsymbol{\zeta}) \colon c(\boldsymbol{s},\lambda,\boldsymbol{\zeta}) \geq 0 \ \forall \boldsymbol{s} \in \tilde{\mathcal{S}}\}, \quad F_2 = \{(\lambda,\boldsymbol{\zeta}) \colon c(\boldsymbol{s},\lambda,\boldsymbol{\zeta}) \geq 0 \ \forall \boldsymbol{s} \in \bar{\mathcal{S}}\}$$

and

$$F_3 = \{(\lambda,\boldsymbol{\zeta}) \colon c(\boldsymbol{s},\lambda,\boldsymbol{\zeta}) \geq 0 \ \forall \boldsymbol{s} \in \mathcal{S}^*\}.$$

A sufficient condition for problem (23) to be equivalent to problem (18) is that $F_1 = F_3$. In order to prove this, we first show that $F_1 = F_2$, and then that $F_2 = F_3$. First, note that $F_1 \subseteq F_2$ as $\bar{\mathcal{S}} \subseteq \tilde{\mathcal{S}}$. On the other hand, we claim that for any $\boldsymbol{s} \in \tilde{\mathcal{S}}$ there exist a $\bar{\boldsymbol{s}} \in \bar{\mathcal{S}}$ such that

$$c(\boldsymbol{s},\lambda,\boldsymbol{\zeta}) \geq c(\bar{\boldsymbol{s}},\lambda,\boldsymbol{\zeta}) \qquad \forall \boldsymbol{\zeta} \geq 0.$$

Note that this claim implies that $F_2 \subseteq F_1$. Indeed, suppose $\boldsymbol{s} = (s_1^1, s_2^1 \ldots, s_{Q+1}^1, \ldots, s_1^N, s_2^N \ldots, s_{Q+1}^N)$ and define $\bar{\boldsymbol{s}}$ as

$$\bar{\boldsymbol{s}} = (s_1^1 + y^1, 0, \ldots, s_{Q+1}^1, s_1^2 + y^2, 0, \ldots, s_{Q+1}^2, \ldots, s_1^N + y^N, 0, \ldots, s_{Q+1}^N)$$

where $y^k = s_2^k + s_3^k + \ldots + s_Q^k$ for any $k \in \mathcal{C}$ (observe that $\bar{\boldsymbol{s}} \in \bar{\mathcal{S}}$). Now,

$$c(\boldsymbol{s}, \lambda, \boldsymbol{\zeta}) = \lambda + \sum_{k\in\mathcal{C}} \Big(\sum_{q=1}^{Q} \mathbb{1}_{\{s_q^k>0\}} a_{q1}^k \mu^k + \sum_{\ell\in\mathcal{C}} s_{Q+1}^{\ell} \sum_{q=1}^{Q} \mathbb{1}_{\{s_q^k>0\}} d_{q1}^{\ell k} \eta^{\ell k} + \sum_{q=1}^{Q} s_q^k g_q^k \theta^k \Big)$$

while

$$c(\bar{\boldsymbol{s}}, \lambda, \boldsymbol{\zeta}) = \lambda + \sum_{k\in\mathcal{C}} \Big(\mathbb{1}_{\{s_1^k+y^k>0\}} a_{11}^k \mu^k + \sum_{\ell\in\mathcal{C}} s_{Q+1}^{\ell} \mathbb{1}_{\{s_1^k+y^k>0\}} d_{11}^{\ell k} \eta^{\ell k} + (s_1^k + y^k) g_1^k \theta^k \Big)$$

From assumption (**A3E**) it follows that $\sum_{q=1}^{Q} \mathbb{1}_{\{s_q^k>0\}} a_{q1}^k \geq \mathbb{1}_{\{s_1^k+y^k>0\}} a_{11}^k$, $\sum_{q=1}^{Q} \mathbb{1}_{\{s_q^k>0\}} d_{q1}^{\ell k} \geq \mathbb{1}_{\{s_q^k+y^k>0\}} d_{11}^{\ell k}$, and $\sum_{q=1}^{Q} s_q^k g_q^k \geq (s_1^k + y^k) g_1^k$, for all $k, \ell \in \mathcal{C}$. As these inequalities for any $\boldsymbol{\zeta} \geq 0$, it follows that $F_2 \subseteq F_1$, and hence $F_1 = F_2$.

Now, we prove that $F_2 = F_3$. On one hand, $F_2 \subseteq F_3$ as $\mathcal{S}^* \subseteq \bar{\mathcal{S}}$. The fact that $F_3 \subseteq F_2$ follows from the following equation (see the proof in Lemma (1) below)

$$c(\bar{\boldsymbol{s}}, \lambda, \boldsymbol{\zeta}) = \sum_{\boldsymbol{s}\in\mathcal{S}^*} c(\boldsymbol{s}, \lambda, \boldsymbol{\zeta}) \prod_{k\in\mathcal{C}} f^k(\bar{\boldsymbol{s}}_1^k, \boldsymbol{s}_1^k) \qquad \forall \lambda \in \mathbb{R}, \boldsymbol{\zeta} \geq 0 \tag{24}$$

where for any $k \in \mathcal{C}$

$$f^k(\bar{\boldsymbol{s}}^k, \boldsymbol{s}^k) = \begin{cases} \frac{\bar{s}_1^k - 1}{n^k - 1}, & \text{if } \bar{s}_1^k > 0 \text{ and } s_1^k = n^k \\ \frac{n^k - \bar{s}_1^k}{n^k - 1}, & \text{if } \bar{s}_1^k > 0 \text{ and } s_1^k = 1 \\ 1, & \text{if } \bar{s}_1^k = 0 \text{ and } s_1^k = 0 \\ 0, & \text{otherwise.} \end{cases} \tag{25}$$

Hence, it can be concluded that $F_1 = F_3$ and therefore programs (23) and (18) are equivalent, as desired.

Lemma 1. *For any $\bar{\boldsymbol{s}} \in \bar{\mathcal{S}}$ and $\boldsymbol{\zeta} \geq 0$ equation (24) holds.*

The proof is by induction on the number of cliques N. Before proceeding note that any $\bar{\boldsymbol{s}} \in \bar{\mathcal{S}}$ is completely characterized by $(\bar{s}_1^1, \bar{s}_1^2, \ldots, \bar{s}_1^N)$, i.e., by the number of nodes in each clique that are in the state $q = 1$ (and the same holds true for $\boldsymbol{s} \in \mathcal{S}^*$ since $\mathcal{S}^* \subseteq \bar{\mathcal{S}}$). We introduce the following notation: For any $k \geq 1$, let $\bar{r}^k = (\bar{s}_1^1, \bar{s}_1^2, \ldots, \bar{s}_1^k)$ and $r^k = (s_1^1, s_1^2, \ldots, s_1^k)$, where $\bar{\boldsymbol{s}} \in \bar{\mathcal{S}}$ and $\boldsymbol{s} \in \mathcal{S}^*$. Additionally, for simplicity, let us denote for any $k \geq 1$,

$$\Pi^k(\bar{r}^k, r^k) = \prod_{\ell=1}^{k} f^{\ell}(\bar{s}_1^{\ell}, s_1^{\ell}), \qquad U^k = \big\{ r^k : r^k = (s_1^1, s_1^2, \ldots, s_1^k),\ s_1^j \in u^j\ \forall j \leq k \big\},$$

where $u^j = \{0, 1, n^j\}$ for any $j \leq k$ (observe that $U^N = \mathcal{S}^*$). In addition, for any $\ell \leq k$ let

$$U^{k\ell} = \{ w^k : w^k = (s_1^1, s_1^2, \ldots, s_1^{\ell-1}, s_1^{\ell+1}, \ldots, s_1^k),\ s_1^j \in \{0, 1, n^j\}\ \forall j \leq k, j \neq \ell \},$$

i.e., $U^{k\ell}$ contain all sequences in U^k but without the state of the ℓ-th clique. Also, define $\Pi^{k\ell}(\bar{w}^k, w^k)$ following the same logic. Before proceeding with the main proof, we prove that for any $\bar{r}^k = (\bar{s}_1^1, \ldots, \bar{s}_1^k)$,

$$\sum_{r^k \in U^k} \Pi^k(\bar{r}^k, r^k)(n^\ell - s_1^\ell) = n^\ell - \bar{s}_1^\ell \qquad \forall \ell \le k, \forall k \ge 1. \tag{26}$$

Pick $\ell \le k$ arbitrarily, then,

$$\begin{aligned}\sum_{r^k \in U^k} \Pi^k(\bar{r}^k, r^k)(n^\ell - s_1^\ell) &= \sum_{w^k \in U^{k\ell}} \sum_{s_1^\ell \in u^\ell} \Pi^{k\ell}(\bar{w}^k, w^k) f^\ell(\bar{s}_1^\ell, s_1^\ell)(n^\ell - s_1^\ell) \\ &= \sum_{w^k \in U^{k\ell}} \Pi^{k\ell}(\bar{w}^k, w^k) \sum_{s_1^\ell \in u^\ell} f^\ell(\bar{s}_1^\ell, s_1^\ell)(n^\ell - s_1^\ell).\end{aligned}$$

We claim $\sum_{s_1^\ell \in u^\ell} f^\ell(\bar{s}_1^\ell, s_1^\ell)(n^\ell - s_1^\ell) = n^\ell - \bar{s}^\ell$. Indeed, if $\bar{s}_1^\ell = 0$ the equation follows from the definition of f^ℓ. Else, then

$$\sum_{s_1^\ell \in u^\ell} f^\ell(\bar{s}_1^\ell, s_1^\ell)(n^\ell - s_1^\ell) = \frac{\bar{s}_1^\ell - 1}{n^\ell - 1}(n^\ell - n^\ell) + \frac{n^\ell - \bar{s}_1^\ell}{n^\ell - 1}(n^\ell - 1) = n^\ell - \bar{s}_1^\ell,$$

and the equation follows. Therefore,

$$\sum_{r^k \in U^k} \Pi^k(\bar{r}^k, r^k)(n^\ell - s_1^\ell) = (n^\ell - \bar{s}_1^\ell) \sum_{w^k \in U^{k\ell}} \Pi^{k\ell}(\bar{w}^k, w^k) = n^\ell - \bar{s}_1^\ell$$

as desired, where the last equation follows as it is readily checked that $\sum_{w^k \in U^{k\ell}} \Pi^{k\ell}(\bar{w}^k, w^k) = 1$.

Now we prove (24) by induction on N. For $N = 1$ the right hand side of equation (24) becomes

$$c(0, \lambda, \boldsymbol{\zeta}) f^1(\bar{s}_1^1, 0) + c(1, \lambda, \boldsymbol{\zeta}) f^1(\bar{s}_1^1, 1) + c(n^1, \lambda, \boldsymbol{\zeta}) f^1(\bar{s}_1^1, n^1).$$

If $\bar{s}_1^1 = 0$ then the result from the definition of f^1 in equation (25). If $\bar{s}_1^1 > 0$ then the above expression becomes

$$\begin{aligned}\lambda + a^1\mu^1 &+ d^{11}\eta^{11}\Big(\frac{n^1 - \bar{s}_1^1}{n^1 - 1}(n^1 - 1) + \frac{\bar{s}_1^1 - 1}{n^1 - 1}(0)\Big) + g^1\theta^1\Big(\frac{n^1 - \bar{s}_1^1}{n^1 - 1}(1) + \frac{\bar{s}_1^1 - 1}{n^1 - 1}(n^1)\Big) \\ &= \lambda + a^1\mu^1 + (n^1 - \bar{s}_1^1) d^{11}\eta^{11} + \bar{s}_1^1 g^1 \theta^1 \\ &= c(\bar{s}_1^1, \lambda, \boldsymbol{\zeta}),\end{aligned}$$

as desired. Now suppose the result holds for $N-1$, we prove it holds for N. First, note that $\bar{\boldsymbol{s}} = \bar{r}^N = (\bar{r}^{N-1}, \bar{s}_1^N)$ for all $\bar{\boldsymbol{s}} \in \bar{\mathcal{S}}$. It is readily seen that

$$c(\bar{r}^N, \lambda, \boldsymbol{\zeta}) = c(\bar{r}^{N-1}, \lambda, \boldsymbol{\zeta}) + c^N(r^N, \boldsymbol{\zeta})$$

where

$$c^N(r^N, \boldsymbol{\zeta}) = 1_{\{\bar{s}_1^N > 0\}}\Big(a^N\mu^N + \sum_{\ell \in \mathcal{C}} (n^\ell - \bar{s}_1^\ell) d^{\ell N} \eta^{\ell N}\Big) + (n^N - \bar{s}_1^N) \sum_{\ell \in \mathcal{C} \setminus \{N\}} d^{N\ell}\eta^{N\ell} + \bar{s}_1^N g^N \theta^N.$$

We have that

$$\sum_{\boldsymbol{s}\in\mathcal{S}^*} c(\boldsymbol{s},\lambda,\boldsymbol{\zeta})\prod_{k\in\mathcal{C}} f^k(\bar{\boldsymbol{s}}_1^k,\boldsymbol{s}_1^k) = \sum_{r^{N-1}\in U^{N-1}}\sum_{s_1^N\in u^N}\big(c(r^{N-1},\lambda,\boldsymbol{\zeta})+c^N(r^N,\boldsymbol{\zeta})\big)\Pi^{N-1}(\bar{r}^{N-1},r^{N-1})f^N(\bar{s}_1^N,s_1^N)$$

$$= \sum_{r^{N-1}\in U^{N-1}}\Big(c(r^{N-1},\lambda,\boldsymbol{\zeta})\Pi^{N-1}(\bar{r}^{N-1},r^{N-1})\sum_{s_1^N\in u^N} f^N(\bar{s}_1^N,s_1^N)+\sum_{s_1^N\in u^N} c^N(r^N,\boldsymbol{\zeta})\Pi^N(\bar{r}^N,r^N)\Big)$$

$$= \sum_{r^{N-1}\in U^{N-1}} c(r^{N-1},\lambda,\boldsymbol{\zeta})\Pi^{N-1}(\bar{r}^{N-1},r^{N-1}) + \sum_{r^{N-1}\in U^{N-1}}\sum_{s_1^N\in u^N} c^N(r^N,\boldsymbol{\zeta})\Pi^N(\bar{r}^N,r^N) \tag{27}$$

$$= c(\bar{r}^{N-1},\lambda,\boldsymbol{\zeta}) + \sum_{r^{N-1}\in U^{N-1}}\sum_{s_1^N\in u^N} c^N(r^N,\boldsymbol{\zeta})\Pi^N(\bar{r}^N,r^N) \tag{28}$$

where equation (27) follows as $\sum_{s_1^N\in u^N} f^N(\bar{s}_1^N,s_1^N)=1$, and equation (28) follows from the induction hypothesis. Therefore, the result follows if and only if

$$c^N(\bar{r}^N,\boldsymbol{\zeta}) = \sum_{r^{N-1}\in U^{N-1}}\sum_{s_1^N\in u^N} c^N(r^N,\boldsymbol{\zeta})\Pi^N(\bar{r}^N,r^N). \tag{29}$$

To prove this equation, first it is readily checked that

$$\sum_{s_1^N\in u^N} c^N(r^N,\boldsymbol{\zeta})f^N(\bar{s}_1^N,s_1^N) = c^N\big((r^{N-1},\bar{s}_1^N),\boldsymbol{\zeta}\big).$$

Hence,

$$\sum_{r^{N-1}\in U^{N-1}}\sum_{s_1^N\in u^N} c^N(r^N,\boldsymbol{\zeta})\Pi^N(\bar{r}^N,r^N) = \sum_{r^{N-1}\in U^{N-1}} c^N\big((r^{N-1},\bar{s}_1^N),\boldsymbol{\zeta}\big)\Pi^{N-1}(\bar{r}^{N-1},r^{N-1})$$
$$= a^N\mu^N + \bar{s}_1^N g^N\theta^N + (n^N-\bar{s}_1^N)\sum_{\ell\in\mathcal{C}} d^{N\ell}\eta^{N\ell} + \sum_{r^{N-1}\in U^{N-1}}\Pi^{N-1}(\bar{r}^{N-1},r^{N-1})\sum_{\ell\in\mathcal{C}\setminus\{N\}}(n^\ell-s_1^\ell)d^{\ell N}\eta^{\ell N}. \tag{30}$$

(where equation (30) follows as $\sum_{r^{N-1}\in U^{N-1}}\Pi^{N-1}(\bar{r}^{N-1},r^{N-1})=1$). Therefore, equation (29) holds if and only if

$$\sum_{\ell\in\mathcal{C}\setminus\{N\}}(n^\ell-\bar{s}_1^\ell)d^{\ell N}\eta^{\ell N} = \sum_{r^{N-1}\in U^{N-1}}\Pi^{N-1}(\bar{r}^{N-1},r^{N-1})\sum_{\ell\in\mathcal{C}\setminus\{N\}}(n^\ell-s_1^\ell)d^{\ell N}\eta^{\ell N}.$$

A sufficient condition for this last equation to hold is that

$$\sum_{r^{N-1}\in U^{N-1}}\Pi^{N-1}(\bar{r}^{N-1},r^{N-1})(n^\ell-s_1^\ell) = n^\ell-\bar{s}_1^\ell \qquad \forall \ell\le N-1,$$

which holds true as (26) holds. Therefore, the inductive step is proven, as desired.

Proof of Proposition 5. From the assumption that there is a directed path from R to all other nodes in $\mathcal{C}$ it follows that $\mathcal{T}$ exist. The proof is divided in two parts: proving primal feasibility, and then proving that the point defined in the statement is a local optimum. For convenience we partition $\mathcal{C}$ as $\mathcal{C} = \mathcal{C}_1 \cup \mathcal{C}_2 \cup \mathcal{C}_3 \cup \mathcal{C}_4$ where

$$\mathcal{C}_1 = \{k \in \mathcal{C} \colon (R,k) \in \mathcal{T}, m^k = c_{\theta^k}/g^k\}, \qquad \mathcal{C}_2 = \{k \in \mathcal{C} \colon (R,k) \in \mathcal{T}, m^k = c_{\mu^k}/a^k\}$$
$$\mathcal{C}_3 = \{k \in \mathcal{C} \colon (R,k) \in \mathcal{T}, m^k = c_{\theta^k}/(n^k g^k) + F(k)\}, \mathcal{C}_4 = \{k \in \mathcal{C} \colon (\ell,k) \in \mathcal{T}, \ell \neq R\}.$$

Observe that the above sets are indeed a partition since $\mathcal{T}$ is a MSA in $\mathcal{G}$. For any $\boldsymbol{s} \in \mathcal{S}^*$ and any value of $\boldsymbol{\zeta} = (\theta^k, \mu^k, \eta^{\ell k})_{k,\ell \in \mathcal{C}}$ define

$$c(\boldsymbol{s}, \boldsymbol{\zeta}) = c(\mathcal{T}) \sum_{k \in \mathcal{C}} \Big\{ g^k \boldsymbol{s}_1^k \theta^k + \mathbb{1}_{\{\boldsymbol{s}_1^k > 0\}} a^k \mu^k + \sum_{\ell \in \mathcal{C}} \boldsymbol{s}_{Q+1}^{\ell} \mathbb{1}_{\{\boldsymbol{s}_1^k > 0\}} d^{\ell k} \eta^{\ell k} \Big\}.$$

Primal feasibility: Let $\boldsymbol{\zeta}^*$ be the value of the decision variables as defined by the statement of the proposition. Note that since $\lambda^* = -1/c(\mathcal{T})$, primal feasibility follows if $c(\boldsymbol{s}, \boldsymbol{\zeta}^*) \geq 1$ for all $\boldsymbol{s} \in \mathcal{S}^*$.

Now, note that after replacing the values $c(\boldsymbol{s}, \boldsymbol{\zeta}^*)$ can be written as

$$c(\boldsymbol{s}, \boldsymbol{\zeta}^*) = \sum_{k \in \mathcal{C}^1} s_1^k + \sum_{k \in \mathcal{C}^2} \mathbb{1}_{\{s_1^k > 0\}} + \sum_{k \in \mathcal{C}_3} \big(\frac{s_1^k}{n^k} + s_{Q+1}^{k'} \frac{\mathbb{1}_{\{s_1^k > 0\}}}{n^{k'}} f(k) \big) + \sum_{k \in \mathcal{C}_4} s_{Q+1}^{\ell(k)} \frac{\mathbb{1}_{\{s_1^k > 0\}}}{n^{\ell(k)}}$$

where we denote by $\ell(k)$ the (unique) node in $\mathcal{G}$ such that $(\ell(k), k) \in \mathcal{T}$ and we define $f(k) = d^{k'k} n^{k'} F(k)/c_{\eta^{k'k}}$. First, observe that if $\boldsymbol{s}$ is such that $s_1^k > 0$ for $k \in \mathcal{C}_1 \cup \mathcal{C}_2$ then $c(\boldsymbol{s}, \boldsymbol{\zeta}^*) \geq 1$. Hence, assume that $\boldsymbol{s}$ satisfies $s_1^k = 0$ for $k \in \mathcal{C}_1 \cup \mathcal{C}_2$, suppose $s_1^k > 0$ for $k \in \mathcal{C}_4$ and that $s_1^k = 0$ for $k \in \mathcal{C}_3$. Since $\mathcal{T}$ is spanning, then there exist a sequence of nodes of $\mathcal{G}$ given by $(l_0, l_1, \cdots, l_y)$, $y \geq 0$, such that $l_0 \in \mathcal{C}_1 \cup \mathcal{C}_2$, $l_z \in \mathcal{C}_4$ for $z \geq 1$, $l_y = k$ and such that $(l_z, l_{z+1}) \in \mathcal{T}$ for all $z = 0, \cdots, y-1$. We claim there exist a $0 \leq z \leq y-1$ such that $s_1^{l_z} = 0$ and $s_1^{l_{z+1}} > 0$. Indeed, since $s_1^{l_0} = 0$ because $l_0 \in \mathcal{C}_1 \cup \mathcal{C}_2$, then the only way the claim would not hold is if $s_1^{l_z} = 0$ for all $0 < z \leq y$; but this is a contradiction with the assumption that $s_1^{\ell_y} > 0$ as $\ell_y = k$. Now, let $0 \leq z \leq y-1$ be such that $s_1^{l_z} = 0$ and $s_1^{l_{z+1}} > 0$. Since $(l_z, l_{z+1}) \in \mathcal{T}$ then $\ell(l_{z+1}) = l_z$. Hence, as $l_{z+1} \in \mathcal{C}_4$ then

$$s_{Q+1}^{\ell(l_{z+1})} \frac{\mathbb{1}_{\{s_1^{l_{z+1}} > 0\}}}{n^{\ell(l_{z+1})}} = n^{l_z} \frac{\mathbb{1}_{\{s_1^{l_{z+1}} > 0\}}}{n^{l_z}} \geq 1,$$

and thus $c(\boldsymbol{s}, \boldsymbol{\zeta}^*) \geq 1$.

Finally, suppose that $\boldsymbol{s}$ is such that $s_1^k = 0$ for all $k \in \mathcal{C}_1 \cup \mathcal{C}_2 \cup \mathcal{C}_4$. Then there exist a $k \in \mathcal{C}_3$ such that $s_1^k > 0$. If $s_1^k = n^k$ then it is evident that $c(\boldsymbol{s}, \boldsymbol{\zeta}^*) \geq 1$. Therefore suppose that $s_1^k = 1$. If $k' = k$ then $s_{Q+1}^k = n^k - 1$, $f(k) = 1$ and thus

$$\frac{s_1^k}{n^k} + s_{Q+1}^{k'} \frac{\mathbb{1}_{\{s_1^k > 0\}}}{n^k} f(k) = \frac{1}{n^k} + \frac{n^k - 1}{n^k} = 1,$$

which implies that $c(\boldsymbol{s},\boldsymbol{\zeta}^*)\geq 1$; hence, suppose that $k'\neq k$. If $s_{Q+1}^{k'}=n^{k'}$, then

$$\frac{s_1^k}{n^k}+s_{Q+1}^{k'}\frac{\mathbb{1}_{\{s_1^k>0\}}}{n^{k'}}f(k)=\frac{1}{n^k}+n^{k'}\frac{1}{n^{k'}}\frac{n^k-1}{n^k}\geq 1,$$

and $c(\boldsymbol{s},\boldsymbol{\zeta}^*)\geq 1$. Thus, suppose that $s_{Q+1}^{k'}<n^{k'}$, and observe this implies that $k'\in\mathcal{C}_3$. If $s_{Q+1}^{k'}=0$, then $s_1^{k'}=n^{k'}$ and since $k'\in\mathcal{C}_3$ it follows that $c(\boldsymbol{s},\boldsymbol{\zeta}^*)\geq$ by a the same argument as above. The other possibility is that $s_{Q+1}^{k'}=n^{k'}-1$, which implies that the sum in $c(\boldsymbol{s},\boldsymbol{\zeta}^*)$ corresponding to $\mathcal{C}_3$ is at least

$$\frac{s_1^k}{n^k}+s_{Q+1}^{k'}\frac{\mathbb{1}_{\{s_1^k>0\}}}{n^k}f(k)+\frac{s_1^{k'}}{n^{k'}}=\frac{1}{n^k}+n^{k'}-1\frac{n^k-1}{n^k}+\frac{1}{n^{k'}}=1+\frac{n^{k'}-1}{n^{k'}n^k},$$

which implies that $c(\boldsymbol{s},\boldsymbol{\zeta}^*)\geq 1$.

From the analysis above it can be concluded that $c(\boldsymbol{s},\boldsymbol{\zeta}^*)\geq 1$ for all $\boldsymbol{s}\in\mathcal{S}^*$ and primal feasibility follows.

Local optimality: We are going to prove that $\boldsymbol{\zeta}^*$ is a local optimum. Global optimality then will follow as the program is linear. Before proceeding with the proof, some additional notation and remarks and necessary. For any $k\in\mathcal{C}_1\cup\mathcal{C}_2\cup\mathcal{C}_3$ let $\mathcal{V}^k$ be the set of nodes of $\mathcal{G}$ that can be reached from k in $\mathcal{T}$ (i.e. the children of node k in $\mathcal{T}$). Formally,

$$\mathcal{V}^k:=\{\ell\in\mathcal{C}:\ \text{there is a directed path from } k \text{ to } \ell \text{ in } \mathcal{T}\}\cup\{k\}.$$

Note we make the convention that $k\in\mathcal{V}^k$.

For any $k\in\mathcal{C}$ we define $\boldsymbol{s}(k),\tilde{\boldsymbol{s}}(k)\in\mathcal{S}^*$ as

$$s_1^\ell(k)=\begin{cases}1, & \text{if } \ell=k\\ 0, & \text{otherwise}\end{cases}\qquad \tilde{s}_1^\ell(k)=\begin{cases}n^k, & \text{if } \ell=k\\ 0, & \text{otherwise.}\end{cases}$$

Observe $\boldsymbol{s}(k)$ ($\tilde{\boldsymbol{s}}(k)$) is the state where all cliques C^ℓ, $\ell\neq k$, have all their nodes recovered and clique C^k has one node (n^k nodes) at state $q=1$, the other nodes being recovered. Similarly, for any $k\in\mathcal{C}_1\cup\mathcal{C}_2\cup\mathcal{C}_3$ define $\tilde{\boldsymbol{s}}(\mathcal{V}^k)$ as

$$\tilde{s}_1^\ell(\mathcal{V}^k)=\begin{cases}n^\ell, & \text{if } \ell\in\mathcal{V}^k\\ 0, & \text{otherwise.}\end{cases}$$

Note that $\tilde{\boldsymbol{s}}(\mathcal{V}^k)$ is the state where all cliques C^ℓ with $\ell\in\mathcal{V}^k$ have all their nodes in state $q=1$, while all the other cliques have all their nodes recovered.

For any $\boldsymbol{s}\in\mathcal{S}^*$ and $\boldsymbol{\zeta}$ define $u(\boldsymbol{s},\boldsymbol{\zeta})=c(\boldsymbol{s},\boldsymbol{\zeta})/c(\mathcal{T})$, that is, $u(\boldsymbol{s},\boldsymbol{\zeta})$ is the left hand side of the constraint of program (18) associated with state $\boldsymbol{s}$ and evaluated at $\boldsymbol{\zeta}$, except for the term λ. Note that for any $k\in\mathcal{C}$,

$$u(\boldsymbol{s}(k),\boldsymbol{\zeta})=g^k\theta^k+a^k\mu^k+(n^k-1)d^{kk}\eta^{kk}+\sum_{\ell\in\mathcal{C},\ell\neq k}n^\ell d^{\ell k}\eta^{\ell k},$$

and

$$u(\tilde{\boldsymbol{s}}(k),\boldsymbol{\zeta}) = n^k g^k \theta^k + a^k \mu^k + \sum_{\ell \in \mathcal{C}, \ell \neq k} n^\ell d^{\ell k} \eta^{\ell k}.$$

Moreover, if $k \in \mathcal{C}_1 \cup \mathcal{C}_2 \cup \mathcal{C}_3$

$$u(\tilde{\boldsymbol{s}}(\mathcal{V}^k),\boldsymbol{\zeta}) = \sum_{\ell \in \mathcal{V}^k} \Big(n^\ell g^\ell \theta^\ell + a^\ell \mu^\ell + \sum_{r \in \mathcal{C} \setminus \mathcal{V}^k} n^r d^{r\ell} \eta^{r\ell} \Big).$$

Remark 6. Suppose that $k \in \mathcal{C}_3$, then it follows that $k' \in \mathcal{V}^k$. Indeed, if that is not the case, consider $\mathcal{T}' = (\mathcal{T} \setminus (R,k)) \cup (k',k)$. As $k' \notin \mathcal{V}^k$, $\mathcal{T}'$ is a spanning arborescence of $\mathcal{G}$. Moreover,

$$\begin{aligned} c(\mathcal{T}') &= c(\mathcal{T}) - \frac{c_{\theta^k}}{n^k g^k} - \frac{c_{\eta^{k'k}}}{n^{k'} d^{k'k}} \frac{n^k - 1}{n^k} + \frac{c_{\eta^{k'k}}}{n^{k'} d^{k'k}} \\ &= c(\mathcal{T}) + \frac{1}{n^k} \Big(\frac{c_{\eta^{k'k}}}{n^{k'} d^{k'k}} - \frac{c_{\theta^k}}{g^k} \Big). \end{aligned}$$

Now, as $k \in \mathcal{C}_3$,

$$\frac{c_{\theta^k}}{n^k g^k} + \frac{c_{\eta^{k'k}}}{n^{k'} d^{k'k}} \frac{n^k - 1}{n^k} < \frac{c_{\theta^k}}{g^k},$$

and from this equation it is readily verified that

$$\frac{c_{\eta^{k'k}}}{n^{k'} d^{k'k}} < \frac{c_{\theta^k}}{g^k}.$$

Hence, $\big(\frac{c_{\eta^{k'k}}}{n^{k'} d^{k'k}} - \frac{c_{\theta^k}}{g^k}\Big) < 0$, and it follows that $c(\mathcal{T}') < c(\mathcal{T})$, which means that $\mathcal{T}'$ is a spanning arborescence (SA) with a smaller cost that $\mathcal{T}$, contradicting the fact that $\mathcal{T}$ is a MSA.$\square$

From the definition of $\boldsymbol{\zeta}^*$ it is readily verified that

$$u(\boldsymbol{s}(k),\boldsymbol{\zeta}^*) = \frac{1}{c(\mathcal{T})} \ \forall k \in \mathcal{C}, \text{ and } u(\tilde{\boldsymbol{s}}(k),\boldsymbol{\zeta}^*) = u(\tilde{\boldsymbol{s}}(\mathcal{V}^k),\boldsymbol{\zeta}^*) = \frac{1}{c(\mathcal{T})} \ \forall k \in \mathcal{C}_2.$$

Moreover, from Remark 6, it also follows that

$$u(\tilde{\boldsymbol{s}}(\mathcal{V}^k),\boldsymbol{\zeta}^*) = \frac{1}{c(\mathcal{T})} \ \forall k \in \mathcal{C}_3.$$

Since $\lambda^* = -1/c(\mathcal{T})$, the previous equations mean that the constraints of program (18) associated with $\boldsymbol{s}(k)$, $k \in \mathcal{C}$, $\tilde{\boldsymbol{s}}(k)$, $k \in \mathcal{C}_2$, and $\tilde{\boldsymbol{s}}(\mathcal{V}^k)$, $k \in \mathcal{C}_2 \cup \mathcal{C}_3$, are tight when evaluated at the point $(\lambda^*, \boldsymbol{\zeta}^*)$. Also, it is readily verified that the cost constraint is also tight at this point.

In order to prove local optimality, we will prove that moving from $\boldsymbol{\zeta}^*$ in any direction to a new point $\hat{\boldsymbol{\zeta}}$ by any amount will either make the program unfeasible or will result in a worse optimal value. A sufficient condition to attain this is to prove that for any $p \in \mathcal{P}$, if ζ_p^* is increased or reduced by any positive amount Δ, then the resulting point is unfeasible or has a worse objective value. For notational simplicity, we will denote by $\hat{\boldsymbol{\zeta}}$ as the resulting point and $\hat{\lambda}$ as the best possible feasible value of λ associated with $\hat{\boldsymbol{\zeta}}$. In particular observe that $\hat{\lambda}$ satisfies

$$-\hat{\lambda} \leq u(\boldsymbol{s}, \hat{\boldsymbol{\zeta}}) \qquad \forall s \in \mathcal{S}^*. \tag{31}$$

Now, let $\mathcal{P}^*$ be the parameters p with $\zeta_p^* > 0$, i.e.,

$$\mathcal{P}^* = \{\theta^k : k \in \mathcal{C}_1 \cup \mathcal{C}_3\} \cup \{\mu^k : k \in \mathcal{C}_2\} \cup \{\eta^{\ell(k)k} : k \in \mathcal{C}_4\} \cup \{\eta^{k'k} : k \in \mathcal{C}_3\}.$$

If ζ_p, $p \in \mathcal{P}^*$ is reduced without increasing the value of other variable, then it is readily verified that $\lambda^* < \hat{\lambda}$ and hence $\hat{\boldsymbol{\zeta}}$ is not optimal. If ζ_p, $p \notin \mathcal{P}^*$ is reduced, then $\hat{\boldsymbol{\zeta}}$ does not satisfy the non-negativity conditions and hence it cannot be feasible.

Therefore, the only remaining possibility is to increase ζ_p, for some $p \in \mathcal{P}$. Observe that since the cost constraint is tight, in order to retain feasibility at least one ζ_p, $p \in \mathcal{P}^*$, must be decreased. In what follows we show that any decrease in any of these ζ_p results in a worse objective value. More precisely, we are going to prove that for any such modification of $\boldsymbol{\zeta}^*$, there always exist a state $\boldsymbol{s} \in \tilde{\mathcal{S}}$ whose constraint is tight at $(\lambda^*, \boldsymbol{\zeta}^*)$ such that

$$u(\boldsymbol{s}, \hat{\boldsymbol{\zeta}}) \leq u(\boldsymbol{s}, \boldsymbol{\zeta}^*).$$

Observe that the above equation implies, from equation (31) and the tightness of $(\lambda^*, \boldsymbol{\zeta}^*)$ at $\boldsymbol{s}$, that $\lambda^* \leq \hat{\lambda}$, and hence $\hat{\boldsymbol{\zeta}}$ does not improve the objective value.

Case I: θ^k, $k \in \mathcal{C}_1$ is reduced by $\Delta > 0$. First, recall that $(\lambda^*, \boldsymbol{\zeta}^*)$ is tight at $\boldsymbol{s}(k)$. Now, if any of θ^ℓ, u^ℓ, $\eta^{r\ell}$, $\ell \neq k$ are increased, then $u(\boldsymbol{s}(k), \hat{\boldsymbol{\zeta}}) = g^k\theta^k - g^k\Delta$, and hence

$$u(\boldsymbol{s}(k), \hat{\boldsymbol{\zeta}}) < g^k\theta^k = u(\boldsymbol{s}(k), \boldsymbol{\zeta}^*).$$

Now, suppose that μ^k is increased. Observe that it can be increased by at most $\Delta(c_{\theta^k}/c_{\mu^k})$. Then

$$u(\boldsymbol{s}(k), \hat{\boldsymbol{\zeta}}) = g^k\theta^k + \frac{\Delta}{c_{\mu^k}}(a^k c_{\theta^k} - g^k c_{\mu^k}).$$

Since $k \in \mathcal{C}_1$, then $c_{\theta^k}/g^k \leq c_{\mu^k}/a^k$, and this implies that $a^k c_{\theta^k} - g^k c_{\mu^k} \leq 0$. Therefore,

$$u(\boldsymbol{s}(k), \hat{\boldsymbol{\zeta}}) \leq g^k\theta^k = u(\boldsymbol{s}(k), \boldsymbol{\zeta}^*),$$

as desired. Suppose that $\eta^{\ell k}$, $\ell \neq k$ is increased. At most it can take the value $\Delta(c_{\theta^k}/c_{\eta^{\ell k}})$. Observe that

$$u(\boldsymbol{s}(k), \hat{\boldsymbol{\zeta}}) = g^k\theta^k + \frac{\Delta}{c_{\eta^{\ell k}}}(n^\ell d^{\ell k} c_{\theta^k} - g^k c_{\eta^{\ell k}}).$$

As $k \in \mathcal{C}_1$, $c_{\theta^k}/g^k \leq c_{\eta^{\ell k}}/d^{\ell k}$ for all $\ell \in \mathcal{C}$, which implies that $(n^\ell d^{\ell k} c_{\theta^k} - g^k c_{\eta^{\ell k}}) \leq 0$ and the result follows. Therefore it can be concluded that if θ^k, $k \in \mathcal{C}_1$ is reduced, increasing any other variables cannot improve the objective value.

In what follows, to simplify the notation and with no loss of generality, it will be assumed that all parameters have unitary costs.

Case II: μ^k, $k \in \mathcal{C}_2$, is reduced by $\Delta > 0$. Recall that $(\lambda^*, \boldsymbol{\zeta}^*)$ is tight at $\boldsymbol{s}(k)$. By the same argument as in the previous case, if any of θ^ℓ, u^ℓ, $\eta^{r\ell}$, $\ell \neq k$ are increased, then $u(\boldsymbol{s}(k), \hat{\boldsymbol{\zeta}}) < a^k \mu^k = u(\boldsymbol{s}(k), \boldsymbol{\zeta}^*)$. If θ^k is increased, then $u(\boldsymbol{s}(k), \hat{\boldsymbol{\zeta}}) = a^k \mu^k + \Delta(g^k - a^k)$, and it follows that $u(\boldsymbol{s}(k), \hat{\boldsymbol{\zeta}}) \leq u(\boldsymbol{s}(k), \boldsymbol{\zeta}^*)$ as $1/a^k \leq 1/g^k$ since $k \in \mathcal{C}_2$.

Suppose that $\eta^{\ell k}$ is increased and that $\ell \notin \mathcal{V}^k$. Then, $u(\boldsymbol{s}(k), \hat{\boldsymbol{\zeta}}) = a^k \mu^k + \Delta(n^\ell d^{\ell k} - a^k)$. If $1/a^k \leq 1/n^\ell d^{\ell k}$, then $(n^\ell d^{\ell k} - a^k) \leq 0$ and the result follows. Hence suppose that $1/(n^\ell d^{\ell k}) < 1/a^k$ and consider $\mathcal{T}' = (\mathcal{T} \setminus (R, k)) \cup (\ell, k)$. Since $\ell \notin \mathcal{V}^k$, then $\mathcal{T}'$ is a SA on $\mathcal{G}$. Moreover, $c(\mathcal{T}') = c(\mathcal{T}) + 1/(n^\ell d^{\ell k}) - 1/a^k < c(\mathcal{T})$, and hence $\mathcal{T}'$ has a lower cost than $\mathcal{T}$, which is a contradiction with the fact that $\mathcal{T}$ is a MSA.

On the other hand, suppose that $\eta^{\ell k}$ is increased and $\ell \in \mathcal{V}^k$, and recall that since $k \in \mathcal{C}_2$ then $\tilde{\boldsymbol{s}}(\mathcal{V}^k)$ is tight at $(\lambda^*, \boldsymbol{\zeta}^*)$. Observe that in this case

$$u(\tilde{\boldsymbol{s}}(\mathcal{V}^k), \hat{\boldsymbol{\zeta}}) = a^k \mu^k - \Delta a^k,$$

which implies that $u(\tilde{\boldsymbol{s}}(\mathcal{V}^k), \hat{\boldsymbol{\zeta}}) < u(\tilde{\boldsymbol{s}}(\mathcal{V}^k), \boldsymbol{\zeta}^*)$, as desired. Therefore, it can be concluded that if μ^k, $k \in \mathcal{C}_2$, is reduced, increasing any other variables cannot improve the objective value.

Case III: θ^k, $k \in \mathcal{C}_3$, is reduced by $\Delta > 0$. Recall that $(\lambda^*, \boldsymbol{\zeta}^*)$ is tight at $\boldsymbol{s}(k)$. By the same argument as in the previous case, if any of θ^ℓ, u^ℓ, $\eta^{r\ell}$, $\ell \neq k$ are increased, then $u(\boldsymbol{s}(k), \hat{\boldsymbol{\zeta}}) < g^k \theta^k + n^{k'} d^{k'k} \eta^{k'k} = u(\boldsymbol{s}(k), \boldsymbol{\zeta}^*)$. Now, suppose that μ^k is increased, and recall that since $k \in \mathcal{C}_3$ then $(\lambda^*, \boldsymbol{\zeta}^*)$ is tight at $\tilde{\boldsymbol{s}}(\mathcal{V}^k)$. Observe that

$$u(\tilde{\boldsymbol{s}}(\mathcal{V}^k), \hat{\boldsymbol{\zeta}}) = n^k g^k \theta^k + \Delta(a^k - n^k g^k).$$

Since $k \in \mathcal{C}_3$, then $1/(n^k g^k) \leq 1/a^k$, which implies that $a^k - n^k g^k \leq 0$ and the desired result follows since $u(\tilde{\boldsymbol{s}}(\mathcal{V}^k), \boldsymbol{\zeta}^*) = n^k g^k \theta^k$.

On the other hand, suppose $\eta^{\ell k}$ is increased, $\ell \neq k'$. If $\ell \in \mathcal{V}^k$ then it is readily noted that $u(\tilde{\boldsymbol{s}}(\mathcal{V}^k), \hat{\boldsymbol{\zeta}}) = n^k g^k \theta^k - \Delta n^k g^k$, and hence $u(\tilde{\boldsymbol{s}}(\mathcal{V}^k), \hat{\boldsymbol{\zeta}}) < u(\tilde{\boldsymbol{s}}(\mathcal{V}^k), \boldsymbol{\zeta}^*)$. Therefore, suppose that $\ell \notin \mathcal{V}^k$. In his case,

$$u(\tilde{\boldsymbol{s}}(\mathcal{V}^k), \hat{\boldsymbol{\zeta}}) = n^k g^k \theta^k + \Delta(n^\ell d^{\ell k} - n^k g^k).$$

If $1/(n^k g^k) \leq 1/(n^\ell d^{\ell k})$ then $(n^\ell d^{\ell k} - n^k g^k) \leq 0$ and the result follows. Hence suppose that $1/(n^\ell d^{\ell k}) < 1/(n^k g^k)$ and consider $\mathcal{T}' = (\mathcal{T} \setminus (R,k)) \cup (\ell,k)$. As $\ell \notin \mathcal{V}^k$ then $\mathcal{T}'$ is a SA on $\mathcal{G}$. Moreover,

$$c(\mathcal{T}') = c(\mathcal{T}) + \frac{1}{n^\ell d^{\ell k}} - \frac{1}{n^k g^k} - \frac{n^k - 1}{n^k n^\ell d^{\ell k}} < c(\mathcal{T}).$$

That is, $\mathcal{T}'$ has a cost strictly less than $\mathcal{T}$, that contradicts the fact that $\mathcal{T}$ is a MSA. Therefore, it can be concluded that if θ^k, $k \in \mathcal{C}_3$, is reduced, increasing any other variables cannot improve the objective value.

Case IV: $\eta^{\ell(k)k}$, $k \in \mathcal{C}_4$, is reduced by Δ, where we denote by $\ell(k)$ the (unique) element of $\mathcal{C}$ such that $(\ell(k),k) \in \mathcal{T}$. As before, recall that $(\lambda^*, \boldsymbol{\zeta}^*)$ is tight at $\boldsymbol{s}(k)$. By the same argument as in the previous case, if any of θ^ℓ, u^ℓ, $\eta^{r\ell}$, $\ell \neq k$ are increased, then $u(\boldsymbol{s}(k), \hat{\boldsymbol{\zeta}}) < n^{\ell(k)} d^{\ell(k)k} \eta^{\ell(k)k} = u(\boldsymbol{s}(k), \boldsymbol{\zeta}^*)$. Suppose θ^k is increased. Then

$$u(\boldsymbol{s}(k), \hat{\boldsymbol{\zeta}}) = n^{\ell(k)} d^{\ell(k)k} \eta^{\ell(k)k} + \Delta(g^k - n^{\ell(k)} d^{\ell(k)k}).$$

If $1/(n^{\ell(k)} d^{\ell(k)k}) \leq 1/g^k$, then $g^k - n^{\ell(k)} d^{\ell(k)k} \leq 0$ and it follows that $u(\boldsymbol{s}(k), \hat{\boldsymbol{\zeta}}) \leq u(\boldsymbol{s}(k), \boldsymbol{\zeta}^*)$. Otherwise, consider $\mathcal{T}' = (\mathcal{T} \setminus (\ell(k),k)) \cup (R,k)$. Then $\mathcal{T}'$ is a SA on $\mathcal{G}$, and $c(\mathcal{T}') = c(\mathcal{T}) + 1/g^k - 1/(n^{\ell(k)} d^{\ell(k)k}) < c(\mathcal{T})$, which contradicts the fact that $\mathcal{T}$ is MSA. Hence the desired result follows.

If μ^k or $\eta^{\ell k}$ are increased, the same argument as above shows that it must be the case that $u(\boldsymbol{s}(k), \hat{\boldsymbol{\zeta}}) \leq u(\boldsymbol{s}(k), \boldsymbol{\zeta}^*)$. Therefore, it can be concluded that if $\eta^{\ell(k)k}$, $k \in \mathcal{C}_4$, is reduced, increasing any other variables cannot improve the objective value.

Case V: $\eta^{k'k}$, $k \in \mathcal{C}_3$ is reduced. Recall that $(\lambda^*, \boldsymbol{\zeta}^*)$ is tight at $\boldsymbol{s}(k)$. By the same argument as in the previous case, if any of θ^ℓ, u^ℓ, $\eta^{r\ell}$, $\ell \neq k$ are increased, then $u(\boldsymbol{s}(k), \hat{\boldsymbol{\zeta}}) < g^k \theta^k + n^{k'} d^{k'k} \eta^{k'k} = u(\boldsymbol{s}(k), \boldsymbol{\zeta}^*)$. Suppose that θ^k is increased. Then

$$u(\boldsymbol{s}(k), \hat{\boldsymbol{\zeta}}) = g^k \theta^k + n^{k'} d^{k'k} \eta^{k'k} + \Delta(g^k - n^{k'} d^{k'k}).$$

Since $k \in \mathcal{C}_3$ it is readily verified that

$$\frac{1}{n^{k'} d^{k'k}} \leq \frac{1}{g^k}, \tag{32}$$

and therefore $g^k - n^{k'} d^{k'k} \leq 0$, which implies that $u(\boldsymbol{s}(k), \hat{\boldsymbol{\zeta}}) \leq u(\boldsymbol{s}(k), \boldsymbol{\zeta}^*)$. If μ^k is increased, then

$$u(\boldsymbol{s}(k), \hat{\boldsymbol{\zeta}}) = g^k \theta^k + n^{k'} d^{k'k} \eta^{k'k} + \Delta(a^k - n^{k'} d^{k'k}).$$

We have that, if $k' = k$, then $a^k - n^k d^{kk} \leq 0$ as $k \in \mathcal{C}_3$. Otherwise, if $k' \neq k$

$$\frac{1}{n^{k'} d^{k'k}} = \frac{1}{n^k} \Big(\frac{1}{n^{k'} d^{k'k}} + \frac{n^k - 1}{n^{k'} d^{k'k}} \Big)$$

$$\leq \frac{1}{n^k}\Big(\frac{1}{g^k} + \frac{n^k - 1}{n^{k'} d^{k'k}}\Big)$$
$$\leq \frac{1}{a^k},$$

where the first inequality follows from equation (32) and the second since $k \in \mathcal{C}_3$. Therefore, $a^k - n^{k'} d^{k'k} \leq 0$ and the result follows.

If $\eta^{\ell k}$ is increased, then

$$u(\boldsymbol{s}(k), \hat{\boldsymbol{\zeta}}) = g^k \theta^k + n^{k'} d^{k'k} \eta^{k'k} + \Delta(n^\ell d^{\ell k} - n^{k'} d^{k'k}),$$

but from the definition of k' it follows that $1/(n^{k'} d^{k'k}) \leq 1/(n^\ell d^{\ell k})$ for all $\ell \in \mathcal{C}$ and therefore $n^\ell d^{\ell k} - n^{k'} d^{k'k} \leq 0$. Thus, it follows that if $\eta^{k'k}$, $k \in \mathcal{C}_3$, is reduced, increasing any other variables cannot improve the objective value.

It can be concluded that by decreasing ζ_p for $p \in \mathcal{P}^*$ and increasing any other variable the objective value of program (18) is not improved. It follows that $\boldsymbol{\zeta}^*$ is a local optimum, as desired.

Proof of Corollary 1. It follows from the closed-form result of Proposition 3.

Proof of Corollary 2. The proof follows from the fact that if $k \notin L_1$ then there exist $\ell \in \mathcal{C} \setminus \{k\}$ (for instance $\ell = \ell(k)$) such that $c_{\eta^{\ell k}}/(n^\ell d^{\ell k}) \leq m^k$. The proof of this fact follows by contradiction: if this is not the case then $\mathcal{T}' = (\mathcal{T} \setminus (\ell(k), k)) \cup (R, k)$ would be a spanning arborescence of $\mathcal{G}$ with a strictly lower cost than $\mathcal{T}$.

Proof of Corollary 3. Follows from a direct application of Proposition 6.

COROLLARY 4. *Suppose the assumptions of Proposition 5 hold and for any $k, \ell \in \mathcal{C}$, $\ell \neq k$, define*

$$x^{\ell k} := \frac{n^\ell d^{\ell k}}{m^\ell n^\ell d^{\ell k} + c_{\eta^{\ell k}}}.$$

Let $k \in \mathcal{C}$ be given and suppose there exist $\ell \in \mathcal{C} \setminus \{k\}$ such that

$$\frac{a^k}{c_{\mu^k}} \leq x^{\ell k} \text{ and } \frac{g^k}{c_{\theta^k}} \leq x^{\ell k},$$

and such that

$$\frac{n^k g^k n^{k'} d^{k'k}}{c_{\theta^k} n^{k'} d^{k'k} + c_{\eta^{k'k}} g^k (n^k - 1)} \leq x^{\ell k}, \text{ when } k' \neq k, \text{ or}$$
$$\frac{n^k g^k d^{kk}}{c_{\theta^k} d^{kk} + c_{\eta^{kk}} g^k} \leq x^{\ell k} \text{ when } k' = k.$$

Then $k \notin L_1$, i.e., it is not optimal to give outside support to clique C^k.

Proof. The result follows from the fact that if $k \in L_1$ then it must hold that $m^\ell + c_{\eta^{\ell k}}/(n^\ell d^{\ell k}) \geq m^k$ for all $\ell \in \mathcal{C} \setminus \{k\}$. Indeed, suppose that there is an $\ell \neq k$ for which it does not hold. If $\ell \notin \mathcal{V}^k$ then $\mathcal{T}' = (\mathcal{T} \setminus (R, k)) \cup (\ell, k)$ would be a spanning arborescence with a lower cost than $\mathcal{T}$. If $\ell \in \mathcal{V}^k$, then $\mathcal{T}' = (\mathcal{T} \setminus ((R, k) \cup (k, \ell))) \cup (R, \ell) \cup (\ell, k)$ would be a spanning arborescence with a lower cost than $\mathcal{T}$. These contradictions yield the desired proof.

Remark 7. Observe that for the above sufficient conditions is not necessary to compute m^ℓ. Indeed, from the definition of m^ℓ the result will still hold if m^ℓ is replaced in the definition of $x^{\ell k}$ by c_{μ^ℓ}/a^ℓ, by c_{θ^ℓ}/g^ℓ, by $c_{\theta^\ell}/(n^\ell g^\ell)+c_{\eta^{r\ell}}(n^\ell-1)/(n^\ell n^r d^{r\ell})$ for any $r \neq \ell$, or by $c_{\theta^\ell}/(n^\ell g^\ell)+c_{\eta^{\ell\ell}}/(n^r d^{r\ell})$.