Balancing Spreads of Influence in a Social Network

Ruben Becker, Federico Corò, Gianlorenzo D'Angelo, Hugo Gilbert
Gran Sasso Science Institute, L'Aquila Italy
ruben.becker@gssi.it

Abstract

The personalization of our news consumption on social media has a tendency to reinforce our pre-existing beliefs instead of balancing our opinions. To tackle this issue, Garimella et al. (NIPS'17) modeled the spread of these viewpoints, also called campaigns, using the independent cascade model introduced by Kempe, Kleinberg and Tardos (KDD'03) and studied an optimization problem that aims to balance information exposure when two opposing campaigns propagate in a network. This paper investigates a natural generalization of this optimization problem in which μ different campaigns propagate in the network and we aim to maximize the expected number of nodes that are reached by at least ν or none of the campaigns, where $\mu \geq \nu \geq 2$. Following Garimella et al., despite this general setting, we also investigate a simplified one, in which campaigns propagate in a correlated manner. While for the simplified setting, we show that the problem can be approximated within a constant factor for any constant μ and ν, for the general setting, we give reductions leading to several approximation hardness results when $\nu \geq 3$. For instance, assuming the gap exponential time hypothesis to hold, we obtain that the problem cannot be approximated within a factor of $n^{-g(n)}$ for any $g(n) = o(1)$ where n is the number of nodes in the network. We complement our hardness results with an $\Omega(n^{-1/2})$-approximation algorithm for the general setting when $\nu = 3$ and μ is arbitrary.

1 Introduction

One of the promises of a highly connected world is that of an impartial spread of opinions driven by free and unbiased sources of information leading to an equitable exposure of opinions to the wide public. On the contrary, the social network platforms that are currently governing news diffusion, while offering many seemingly-desirable features like searching, personalization, and recommendation, are reinforcing the centralization of information spreading and the creation of what is often termed *echo chambers* and *filter bubbles* (Garimella et al. 2018). Stated differently, algorithmic personalization of news diffusion are likely to create homogeneous polarized clusters where users get less exposure to conflicting viewpoints. A good illustration of this issue was given by Conover et al. (2011) who studied the Twitter network during the 2010 US congressional midterm elections. The authors demonstrated that the retweet network had a highly segregated partisan structure with extremely limited connectivity between left-wing and right-wing users. More recently, a similar finding has been obtained by the Electome project at the MIT Media Lab for the 2016 US presidential elections. Indeed, several illustrations issued from this research project and reported by Thompson (2016) clearly demonstrate that the Twitter network (at the time of the study) contained two clusters consisting of supporters of D. Trump and H. Clinton, respectively, that had a strong inner connectivity while being poorly connected to one another.

Consequently, instead of giving users a diverse perspective and balancing users opinions by exposing them to challenging diverse ideas, social media platforms are likely to make users more extreme by only exposing them to views that reinforce their pre-existing beliefs (Conover et al. 2011; Del Vicario et al. 2016).

To address this issue from an algorithmic perspective, Garimella et al. (2017) introduced the problem of *balancing information exposure* in a social network. Following the *influence maximization paradigm* going back to the seminal work of Kempe, Kleinberg, and Tardos (2003; 2005; 2015), their problem involves two opposing viewpoints or campaigns that propagate in a social network following the *independent cascade model*. Given initial seed sets for both campaigns, they consider the optimization problem of selecting at most k additional seed nodes for both campaigns in order to maximize the expected number of nodes that are reached by either both or none of the campaigns. The authors studied two different settings, namely the *heterogeneous* and *correlated* settings. The heterogeneous setting corresponds to the general case in which there is no restriction on the probabilities with which the campaigns spread. Contrarily, in the correlated setting, the probability distributions for different campaigns are identical and completely correlated. After proving the NP-hardness of balancing information exposure, the authors designed efficient approximation algorithms with an approximation ratio of $(1 - 1/e - \epsilon)/2$ for any $\epsilon > 0$ for both the correlated and heterogeneous settings.

Our Contribution. In this work, we address the main open problem in the work of Garimella et al. (2017). That is, we generalize their optimization problem to a setting with arbitrarily many campaigns. More precisely, let μ and ν be fixed constants such that $2 \leq \nu \leq \mu$. Assume μ opposing campaigns are spreading in a social network and the task is to maximize the number of users that are reached by at least ν campaigns or remain oblivious to all of them. We term this problem the μ-ν-BALANCE problem. This generalization is motivated by the fact that for most problems, not only two, but a multitude of viewpoints are perceivable. This may simply be due to the complexity of the problems or due to the wide diversity of sensibilities present in our modern societies. Hence, given this possibly large number μ of viewpoints, the ν threshold parameter aims to guarantee that influenced users are exposed to a sufficiently large subset of viewpoints, hopefully providing them with a more representative picture.

Interestingly, we obtain results that surprisingly differ from the ones Garimella et al. (2017) obtained for the special case where $\mu = \nu = 2$. Indeed, while we show in Section 5 that any μ-ν-BALANCE problem can be approximated within a constant factor in the correlated setting, in Section 3, we obtain strong approximation hardness results for the heterogeneous setting. In particular, when $\nu \geq 3$, we show that under the Gap Exponential Time Hypothesis (Gap-ETH) (Manurangsi 2017), there is no $n^{-g(n)}$-approximation algorithm with $g(n) = o(1)$ for the μ-ν-BALANCE problem, where n is the number of nodes. Moreover, when $\nu \geq 4$, we show that if a certain class of one-way functions exists (a common assumption in the field of cryptography (Applebaum 2013)), there is no $n^{-\epsilon}$-approximation algorithm for the μ-ν-BALANCE problem, where $\epsilon > 0$ is a constant which depends on ν. We mitigate these hardness results in Section 4 by designing an algorithm with an approximation factor of $\Omega(n^{-1/2})$ for the case where $\nu = 3$ and μ is an arbitrary constant.

Detailed proofs for all results discussed in this paper are included in the extended version of this article that is available online (Becker et al. 2019).

Related work. There is a vast literature on influence maximization and the problem has been studied by various communities. We refer the interested reader to (Borgs et al. 2014; Kempe, Kleinberg, and Tardos 2015) and references therein for an algorithmic account of this field. Recently, the influence maximization paradigm has raised some ethical challenges. These challenges have very diverse objectives as limiting the spread of a "bad" campaign by starting the spreading of a "good" campaign blocking the first one (Budak, Agrawal, and El Abbadi 2011), maximizing social welfare by controlling the spread of multiple campaigns and orienting them towards the right agents (Borodin et al. 2017) or minimizing the access gap between the highly connected individuals and the poorly connected ones (Fish et al. 2019). We here focus on works that tackle algorithmically the challenges related to diversity.

Two works closely related to ours are the ones of Aslay et al. (2018) and Matakos and Gionis (2018). Indeed, both of these works aim to break filter bubbles. The former work tackles an item-aware information propagation problem in which a centralized agent must recommend some articles to a small set of seed users such that the spread of these articles maximizes the expected diversity of exposure of agents. The diversity of exposure is measured by a sum of agent-dependent functions that takes into account user leanings. The authors show that the NP-hard problem they define amounts to optimizing a monotone and submodular function under a matroid constraint and design a constant factor approximation algorithm. The latter paper models the problem of maximizing the diversity of exposure in a social network as a quadratic knapsack problem. Here also the problem amounts to recommending a set of articles to some users in order to maximize a diversity index taking into account users' leanings and the strength of their connections in the network. The authors show that the resulting diversity maximization problem is inapproximable and design a polynomial algorithm without approximation guarantee.

Conversely, instead of maximizing the diversity of information a user gets, one could wish to maximize the diversity of users a campaign reaches. This problem, was tackled by Tang et al. (2014). The authors presented several optimization objectives to take into account user-diversity and gave empirical evidence that greedy algorithms perform well in practice for optimizing them. A closely related work is the one of Tsang et al. (2019). The authors introduced the problem of maximizing the spread of a campaign while respecting a group-fairness constraint. In their setting, each user of the network belongs to one or several communities and the authors defined several criteria to guarantee that each community gets its fair share of information. For each of these criteria, they showed that maximizing influence while respecting the related fairness constraint can be casted as a multi-objective optimization problem. Lastly, they designed an algorithm to tackle such multi-objective influence problems that provides an asymptotic approximation guarantee of $1 - 1/e$.

2 Preliminaries

For $n \in \mathbb{N}$, $[n]$ denotes the set $\{1, \dots, n\}$. We generalize set operators to sequences of sets: For two sequences of sets $\mathcal{A}$, $\mathcal{A}'$, both of size μ, and a set A, we let $\mathcal{A} \cup \mathcal{A}' = (A_i \cup A'_i)_{i \in [\mu]}$ be the sequence of element-wise unions and $\mathcal{A} \cap A = (A_i \cap A)_{i \in [\mu]}$ be the sequence of element-wise intersections with the set A.

Independent Cascade model. The well-known *Independent Cascade Model* (ICM) introduced by Kempe, Kleinberg, and Tardos (2015) is one of the best studied models for information spread in social networks. Given a directed graph $G = (V, E)$, probabilities $p : E \to [0, 1]$ and an initial node set $A \subseteq V$ called *seed nodes*, define $A_0 = A$, and, for $t \geq 0$, call a node $v \in A_t$ *active* at time t. A node active at time t remains active at latter time steps. If v is active at time $t \geq 0$ but was not active at time $t-1$, i.e., $v \in A_t \setminus A_{t-1}$ (formally let $A_{-1} = \emptyset$), it tries to activate each neighbor w, independently, and succeeds with probability p_{vw}. In case of success w becomes active at step $t+1$, i.e., $w \in A_{t+1}$. If at some time $t^* \geq 0$, we have that $A_{t^*} = A_{t^*+1}$ we say that the process has *quiesced* and call t^* the time of quiescence. For an initial set A, $\sigma(A) = \mathrm{E}[|A_{t^*}|]$ denotes the expected number of nodes activated at the time of quiescence when running the process with seed nodes A. Kempe, Kleinberg, and Tardos showed that this process is a special case of what is referred to as the *Triggering Model*, see (Kempe, Kleinberg, and Tardos 2015, Proof of Theorem 4.5). For a node $v \in V$, let N_v denote all in-neighbors of v. Here, every node independently picks a *triggering set* $T_v \subseteq N_v$ according to a distribution over subsets of its in-neighbors, namely $T_v = S$ with probability $\prod_{u \in S} p_{uv} \cdot \prod_{u \in N_v \setminus S}(1-p_{uv})$. For a possible outcome $X = (T_v)_{v \in V}$ of triggering sets for the nodes V, let $\rho_X(A)$ be the set of nodes reachable from A in the outcome X. Note that after sampling X, the quantity $\rho_X(A)$ is deterministic. According to Kempe, Kleinberg, and Tardos (2015), this model is equivalent to the ICM and it holds that $\sigma(A) = \mathrm{E}_X[|\rho_X(A)|]$. While it is not feasible to compute $\rho_X(A)$ for all outcome profiles X, a $(1 \pm \epsilon)$-approximation to $\sigma(A)$ can be obtained with probability at least $1-\delta$ by sampling $\Omega(|V|^2 \log(1/\delta)/\epsilon^2)$ possible outcomes X and computing the average over the corresponding values $|\rho_X(A)|$, see (Kempe, Kleinberg, and Tardos 2015, Proposition 4.1).

The μ-ν-Balance problem

Inspired by Garimella et al. (2017), we consider several information spread processes, also called "campaigns", unfolding in parallel, each following the ICM described above. Formally, we are given a graph $G = (V, E)$ and μ probability functions $(p_i)_{i \in [\mu]}$, where each p_i is a probability function as in the ICM described above, i.e., $p_i : E \to [0, 1]$. For an index $i \in [\mu]$, let $X_i = (T_v)_{v \in V}$ be a possible outcome sampled using probabilities p_i. Then for a seed set $A \subseteq V$, we denote with $\rho^{(i)}_{X_i}(A)$ the set of nodes reachable from A in outcome X_i. Now, let $\mathcal{X} = (X_i)_{i \in [\mu]}$ be an outcome profile by letting each X_i be a possible outcome according to distribution p_i. Then, for $\mathcal{A} = (A_i)_{i \in [\mu]}$ with $A_i \subseteq V$, we denote with

$$\rho_{\mathcal{X}}(\mathcal{A}) = (\rho^{(i)}_{X_i}(A_i))_{i \in [\mu]}$$

the set of reached nodes in outcome profile $\mathcal{X}$ from seed sets $\mathcal{A}$. Last, for a sequence $\mathcal{R}$=$(R_i)_{i \in [\mu]}$ of subsets of V,

$$\mathrm{NoSM}_{\mu,\nu}(\mathcal{R}) := \left|(V \setminus \bigcup_{i \in [\mu]} R_i) \cup \bigcup_{M \subseteq [\mu] : |M| \geq \nu} \bigcap_{i \in M} R_i\right|$$

is defined to be the number of nodes that are contained in **N**one **o**r **S**ufficiently **M**any, i.e., at least ν, of the sets in $\mathcal{R}$. The $\mathrm{NoSM}_{\mu,\nu}$-function allows us to formalize our objective function of maximizing the number of nodes that are reached by none of sufficiently many of the campaigns.[1]

[1]In the special case studied by Garimella et al. (2017), the objective function is modelled by a set difference operator. Sadly, in the general case, such a straightforward formulation is not conceivable. We resolve this issue by introducing the $\mathrm{NoSM}_{\mu,\nu}$-function.

Problem statement. For constant integers $\mu \geq \nu \geq 2$, we define the μ-ν-Balance problem as follows.

μ-ν-Balance

Input: Graph $G = (V, E)$, probabilities $\mathcal{P} = (p_i)_{i \in [\mu]}$, seed sets $\mathcal{I} = (I_i)_{i \in [\mu]}$, and an integer k.

Find: Sets $\mathcal{S} = (S_i)_{i \in [\mu]}$ with $\sum_{i \in [\mu]} |S_i| \leq k$, such that $\Phi^{\mathcal{I}}_{\mu,\nu}(\mathcal{S})$ is maximum, where

$$\Phi^{\mathcal{I}}_{\mu,\nu}(\mathcal{S}) := \mathrm{E}_{\mathcal{X}}[\mathrm{NoSM}_{\mu,\nu}(\rho_{\mathcal{X}}(\mathcal{I} \cup \mathcal{S}))].$$

We refer to the objective function simply by $\Phi(\mathcal{S})$, in case $\mathcal{I}$, μ, and ν are clear from the context. We assume $k \leq \nu|V|$ as otherwise the problem becomes trivial by choosing $S_i = V$ for every $i \in [\nu]$. Moreover, we assume w.l.o.g. that $|V| \geq \mu$ and $k \geq \nu$, since μ and ν are constant numbers and the μ-ν-Balance problem becomes computationally easy if $|V|$ or k is a constant (one could just try all feasible solutions). Following Garimella et al. (2017), we distinguish two settings. (1) The *heterogeneous* setting corresponds to the general case in which there is no restriction on $\mathcal{P}$. (2) In the *correlated* setting, the distributions p_i are identical and completely correlated for all $i \in [\mu]$. That is, if an edge (u, v) propagates a campaign to v, it propagates all campaigns that reach u to v.

Decomposing the Objective Function. In all of our algorithms, we use the approach of decomposing the objective function into summands and approximating the summands separately. For an outcome profile $\mathcal{X}$, and seed sets $\mathcal{I} = (I_i)_{i \in [\mu]}$, we define $V^{\ell,\mathcal{I}}_{\mathcal{X}} \subseteq V$, for $\ell = 0, \ldots, \mu$, to be the set of nodes that are reached by exactly ℓ campaigns *from the seed sets* $\mathcal{I}$. Formally, for any value $\ell \in [\mu]$,

$$V^{\ell,\mathcal{I}}_{\mathcal{X}} := \bigcup_{\tau \in \binom{[\mu]}{\ell}} \Big(\bigcap_{i \in \tau} \rho^{(i)}_{X_i}(I_i) \setminus \bigcup_{j \in [\mu] \setminus \tau} \rho^{(j)}_{X_j}(I_j) \Big),$$

where $\binom{[\mu]}{\ell} = \{\tau \subseteq [\mu] \ : \ |\tau| = \ell\}$. We write $V^{\ell}_{\mathcal{X}}$, if the initial seed sets $\mathcal{I}$ are clear from the context. In the above definition, by convention an empty union is the empty set, while an empty intersection is the whole universe, here V. Accordingly, we define

$$\Phi^{\ell}(\mathcal{S}) := \mathrm{E}_{\mathcal{X}}[\mathrm{NoSM}_{\mu,\nu}(\rho_{\mathcal{X}}(\mathcal{I} \cup \mathcal{S}) \cap V^{\ell,\mathcal{I}}_{\mathcal{X}})].$$

Note that $\Phi^{\ell}(\mathcal{S})$ measures the expected number of nodes that are reached by 0 or at least ν campaigns among nodes that have been reached by exactly ℓ campaigns from $\mathcal{I}$. Now, the objective function decomposes as

$$\Phi(\mathcal{S}) = \mathrm{E}_{\mathcal{X}}[\mathrm{NoSM}_{\mu,\nu}(\rho_{\mathcal{X}}(\mathcal{I} \cup \mathcal{S}))] =$$
$$\mathrm{E}_{\mathcal{X}}\Big[\sum_{\ell \in [\mu]} \mathrm{NoSM}_{\mu,\nu}(\rho_{\mathcal{X}}(\mathcal{I} \cup \mathcal{S}) \cap V^{\ell}_{\mathcal{X}})\Big] = \sum_{\ell \in [\mu]} \Phi^{\ell}(\mathcal{S}),$$

using linearity of expectation and that sets $V^{\ell}_{\mathcal{X}}$ are disjoint. Furthermore, we will denote by

$$\Phi^{\geq \ell}(\mathcal{S}) := \mathrm{E}_{\mathcal{X}}[\mathrm{NoSM}_{\mu,\nu}(\rho_{\mathcal{X}}(\mathcal{I} \cup \mathcal{S}) \setminus (\cup_{j=0}^{\ell-1} V^{j}_{\mathcal{X}}))].$$

the expected number of nodes that are reached by sufficiently many campaigns or none of them among the nodes that have previously been reached by *at least* ℓ campaigns. Clearly, $\Phi^{\geq \ell}(\mathcal{S}) = \sum_{i=\ell}^{\mu} \Phi^i(\mathcal{S})$ and $\Phi^{\geq 0}(\mathcal{S}) = \Phi(\mathcal{S})$. For convenience, in what follows, we will often refer to $\mathcal{S}$ as a set of pairs in $\hat{V} := V \times [\mu]$, where picking pair (v, i) into $\mathcal{S}$ corresponds to picking v into set S_i. We fix the following observations:

- For $\ell = 0$, $\Phi^0(\mathcal{S})$ is maximized by $\mathcal{S} = (\emptyset)_{i\in[\mu]}$. The achieved value is the expected size of $V^0_{\mathcal{X}}$: $\Phi^0(\mathcal{S}) = \mathrm{E}_{\mathcal{X}}[\mathrm{NoSM}_{\mu,\nu}(\rho_{\mathcal{X}}(\mathcal{I} \cup (\emptyset)_{i\in[\mu]}) \cap V^0_{\mathcal{X}})] = \mathrm{E}_{\mathcal{X}}[|V^0_{\mathcal{X}}|]$.
- For $\ell = \nu - 1$, the function $\Phi^{\geq \nu-1}(\mathcal{S}) = \sum_{i=\nu-1}^{\mu} \Phi^i(\mathcal{S})$ is monotone and submodular.

The First Structural Lemma. When applying the greedy hill climbing algorithm to find a set of size k maximizing a submodular set function, the key property that is used in the analysis is that, at any stage, there exists an element that leads to an improvement that is at least a fraction of k of the difference of the optimal and the current solution, compare for example (Hochbaum 1997, Lemma 3.13). Maybe the most important structural lemma underlying our algorithms is a very similar result for the functions $\Phi^{\geq \ell}$.

Lemma 1. *Let $\ell \in [1, \nu - 1]$ and $\mathcal{S} \subseteq \hat{V}$ with $|\mathcal{S}| \leq k - (\nu - \ell)$ and define $U := \{\tau \subseteq \hat{V}, |\tau| = \nu - \ell\}$. Then, $\tau^* = \arg\max\{\Phi^{\geq \ell}(\mathcal{S} \cup \tau) : \tau \in U\}$ satisfies*

$$\Phi^{\geq \ell}(\mathcal{S} \cup \tau^*) - \Phi^{\geq \ell}(\mathcal{S}) \geq \frac{\Phi^{\geq \ell}(\mathcal{S}^*_{\geq \ell}) - \Phi^{\geq \ell}(\mathcal{S})}{\binom{k}{\nu-\ell}},$$

where $\mathcal{S}^*_{\geq \ell}$ is a solution of size k maximizing $\Phi^{\geq \ell}$.

The Correlated Case. For the correlated setting, where probability functions are identical for all campaigns and the cascade processes are completely correlated, we introduce an additional function called Ψ. First note that in this setting, the outcome profile $\mathcal{X}$ in the definition of $\Phi(\mathcal{S})$ satisfies $X_1 = \ldots = X_\mu$. In order to define Ψ, we introduce an additional fictitious campaign, call it campaign 0, that spreads with the same probability $p_0 = p_1 = \ldots = p_\mu$ as the other μ campaigns. We extend the outcome $\mathcal{X} = (X_i)_{i\in[\mu]}$ with $X_1 = \ldots = X_\mu$ to contain also an identical copy X_0 and define $\Psi : 2^{V\times\{0\}} \to [n]$ by

$$\Psi(\mathcal{T}) := \mathrm{E}_{\mathcal{X}}\left[\left|\left(\rho^{(0)}_{X_0}(\mathcal{T}) \cap \bigcup_{j=1}^{\nu-1} V^j_{\mathcal{X}}\right) \cup \bigcup_{j=\nu}^{\mu} V^j_{\mathcal{X}}\right|\right].$$

We will explain the rationale behind Ψ in Section 5. For now, it is only important to observe that Ψ is monotone and submodular in $\mathcal{T}$, following from σ having these properties.

Approximating Ψ and $\Phi^{\geq \ell}$. As mentioned above, already in the standard ICM, it is not feasible to evaluate the function σ exactly. However, σ can be approximated to within a factor of $(1 \pm \epsilon)$ by sampling a polynomial number of times. A very similar approach works for approximating the functions Ψ and $\Phi^{\geq \ell}$ for $\ell \in [0, \nu]$. That is, there is an algorithm $\mathtt{approx}(f, \mathcal{S}, \mathcal{I}, \nu, \epsilon, \delta)$ that, for $f \in \{\Psi, \Phi^{\geq 0}, \ldots, \Phi^{\geq \nu}\}$, sets $\mathcal{S}$ and $\mathcal{I}$, and parameters ν, ϵ, δ returns a $(1 \pm \epsilon)$-approximation of $f(S)$ with probability $1 - \delta$.

Maximizing $\Phi^{\geq \nu-1}$ and Ψ. Here, we fix the result that the standard greedy hill climbing algorithm, we refer to it as $\textsc{Greedy}(f, \epsilon, \delta, \mathcal{I}, \nu, k)$, can be applied in order to approximate both $f \in \{\Phi^{\geq \nu-1}, \Psi\}$ to within a factor of $1 - 1/e - \epsilon$ for any $0 < \epsilon < 1$ with probability at least $1 - \delta$ for any $0 < \delta \leq 1/2$. This is based on the fact that these functions are submodular and monotone set functions. Since we can only evaluate $\Phi^{\geq \nu-1}$ and Ψ approximately, we obtain the additive ϵ-term.

Lemma 2. *Let $f \in \{\Phi^{\geq \nu-1}, \Psi\}$ and let $0 < \epsilon < 1$ and $0 < \delta \leq 1/2$. With probability at least $1 - \delta$, $\textsc{Greedy}(f, \epsilon, \delta, \mathcal{I}, \nu, k)$ returns $\mathcal{S}$ satisfying*

$$f(\mathcal{S}) \geq \left(1 - \frac{1}{e} - \epsilon\right) \cdot f(\mathcal{S}^*),$$

where $\mathcal{S}^*$ is an optimal solution of size k to maximizing f.

3 Hardness of Approximation for the Heterogeneous Case

In this section, we show that in the heterogeneous setting, the μ-ν-BALANCE problem is as hard to approximate as the DENSEST-k-SUB-d-HYPERGRAPH problem (Chlamtác et al. 2018) for $\nu \geq d + 1$. This result even holds for the deterministic variant of the μ-ν-BALANCE problem in which all edge probabilities are equal to 0 or 1.

We start by recalling the DENSEST-k-SUB-d-HYPERGRAPH problem which is defined on d-regular hypergraphs, i.e., hypergraphs in which all hyperedges are composed of exactly d vertices.

DENSEST-k-SUB-d-HYPERGRAPH
Input: d-Regular Hypergraph $G = (V, E)$, integer k
Find: Set $S \subseteq V$ with $\lvert S\rvert \leq k$, s.t. $\lvert E(S)\rvert$ is maximal, where $E(S) := \{e \in E : e \subseteq S\}$.

In particular, when d equals 2, DENSEST-k-SUB-d-HYPERGRAPH is known as the more classical DENSEST-k-SUBGRAPH problem.

Our hardness of approximation results are summarized in the following theorem.

Theorem 1. *Let $d \geq 2$ and $\mu \geq \nu \geq d+1$. Define $p = d!/d^d$ and $\lambda = d!\binom{\mu-\nu+d}{d}$, and consider the following two cases:*

Case $d = 2$: *Let $\alpha(n) = n^{-g(n)}$ with g being non increasing*[2]*, $g(n) = o(1)$, $\alpha(n) \in (0, 1]$ and $\beta(n) = \frac{p\cdot n^{-6g(n)}}{2\lambda}$.*

Case $d \geq 3$: *Let $\alpha(n) = n^{-\epsilon(d)}$ where $\epsilon(d) > 0$ is a constant which depends on d, $\alpha(n) \in (0, 1]$ and $\beta(n) = \frac{p\cdot n^{-\epsilon'(d)}}{2\lambda}$, with $\epsilon'(d) = (2d + 2) \cdot \epsilon(d)$.*

*In both cases, the following statement holds: if there is an $\alpha(|\overline{V}|)$-approximation algorithm for the deterministic μ-ν-*BALANCE *problem, then there is a $\beta(|V|)$-approximation algorithm for* DENSEST-k-SUB-d-HYPERGRAPH. *Here, $|\overline{V}|$ (resp. $|V|$) is the number of vertices in the μ-ν-*BALANCE *(resp.* DENSEST-k-SUB-d-HYPERGRAPH*) problem.*

[2]Note that the assumption that g is non-increasing can be made w.l.o.g. for this result.

We proceed by describing some notable consequences of the above theorem:

1. As DENSEST-k-SUB-d-HYPERGRAPH with $d \geq 3$ cannot be approximated within $1/n^{\epsilon}$ for some constant $\epsilon > 0$ depending on d if a particular class of one way functions exists (Applebaum 2013), the same hardness result holds for any μ-ν-BALANCE problem with $\nu \geq d+1 \geq 4$.
2. Since DENSEST-k-SUBGRAPH cannot be approximated within $1/n^{o(1)}$ if the Gap Exponential Time Hypothesis (Gap-ETH) holds (Manurangsi 2017), the same hardness result holds for any μ-ν-BALANCE problem with $\nu \geq 3$.

Further (conditional) approximation hardness results are known for DENSEST-k-SUBGRAPH and these approximation hardness results also transfer to the μ-ν-BALANCE problem with $\nu \geq 3$. We review some of these results: DENSEST-k-SUBGRAPH cannot be approximated within any constant if the unique games with small set expansion conjecture holds (Raghavendra and Steurer 2010). It cannot be approximated within $n^{-(\log \log n)^{-c}}$ for some constant c if the exponential time hypothesis holds (Manurangsi 2017). It is easy to see that the reduction that we design in this section allows to transfer these hardness results to μ-ν-BALANCE with $\nu \geq 3$.

The Reduction. We now detail the reduction on which Theorem 1 builds. We consider the following transform τ of an instance $(G=(V,E),k)$ of the DENSEST-k-SUB-d-HYPERGRAPH problem into an instance

$$\tau(G,k) = (\overline{G} = (\overline{V}, \overline{E}), \mathcal{P}, \mathcal{I}, \overline{k})$$

of the μ-ν-BALANCE problem.

- Define $\overline{V} := V_{\square} \cup V_{\bigcirc}$, where $V_{\square} := V$, i.e., for each node $v \in V$, we get a node v in $\overline{V}$. We refer to $V_{\bigcirc}$ as the circle-nodes and to $V_{\square}$ as the box-nodes. Moreover, we let $J := \binom{[\mu-\nu+d]}{d}$, and S_d be the set of permutations of $[d]$; we then define $V_{\bigcirc}$ as
$$V_{\bigcirc} := \{e^t_{\iota,\pi} : e \in E, \iota \in J, \pi \in S_d, t \in [l]\},$$
i.e., for each edge $e \in E$, we create λl circle-nodes, where $l := |V|+1$ and $\lambda := |S_d| \cdot |J| = d!\binom{\mu-\nu+d}{d}$. That is, each set ι of d campaigns in J, induces l circle-nodes $e^t_{\iota,\pi}, t \in [l]$ for each permutation π in S_d.
- The arc set $\overline{E}$ and the probabilities are defined as shown in Figure 1 illustrating the case $d = 3$. We get this scheme in $\overline{G}$ for every edge $\{v_1, \ldots, v_d\} \in E$, for each permutation π in S_d, and for each set in J of d campaigns.
- The initial seed sets $\mathcal{I}$ are defined as
$$I_1 = I_2 = \ldots = I_{\mu-\nu+d} = \emptyset, \quad I_{\mu-\nu+d+1} = \ldots = I_{\mu} = \overline{V}.$$
- The budget is the same as in the DENSEST-k-SUB-d-HYPERGRAPH problem, i.e., $\overline{k} = k$.

Note that each node in $\overline{G}$ is already covered by $\nu - d$ campaigns and that the instance generated is deterministic, in the sense that probability values are either 0 or 1.

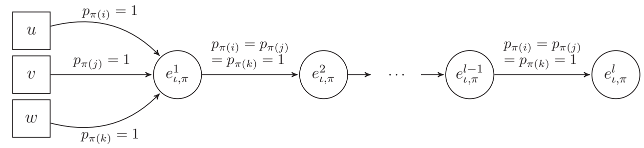

Figure 1: This figure illustrates the case $d = 3$. For an hyperedge $e = \{u, v, w\}$ in G, we get $d!\binom{\mu-\nu+d}{d}$ schemes of the above type, one for each set $\iota = \{i, j, k\} \in J$ and for each way of ordering them given by a permutation $\pi \in S_d$. Probabilities that are not given are equal to 0.

Let us now fix a μ-ν-BALANCE instance $P = (\overline{G} = (\overline{V}, \overline{E}), \mathcal{P}, \mathcal{I}, \overline{k})$ resulting from the transform τ as image of a DENSEST-k-SUB-d-HYPERGRAPH instance $Q = (G = (V,E),k)$. Clearly, $\overline{V}$ is of cardinality $|V| + \lambda l|E|$ and $\overline{E}$ is of cardinality $\lambda(l+d-1)|E|$. Let us denote by Σ the set of feasible solutions for P. For each $\mathcal{S} \in \Sigma$, it holds that the objective function $\Phi(\mathcal{S})$ can be decomposed as

$$\begin{aligned}\Phi(\mathcal{S}) = \Phi_{\square}(\mathcal{S}) + \Phi_{\bigcirc}(\mathcal{S}),&\\ \text{where } \Phi_{\square}(\mathcal{S}) := \mathrm{NoSM}_{\mu,\nu}(\rho_{\mathcal{X}}(\mathcal{I} \cup \mathcal{S}) \cap V_{\square})&\\ \text{and } \Phi_{\bigcirc}(\mathcal{S}) := \mathrm{NoSM}_{\mu,\nu}(\rho_{\mathcal{X}}(\mathcal{I} \cup \mathcal{S}) \cap V_{\bigcirc}),&\end{aligned}$$

for $\mathcal{X}$ being the only possible (deterministic) outcome profile. Now, let $\mathcal{S}^*$, $\mathcal{S}^*_{\square}$, and $\mathcal{S}^*_{\bigcirc}$ denote optimal solutions to the problem of maximizing Φ, $\Phi_{\square}$, and $\Phi_{\bigcirc}$, respectively, over Σ. The following lemma collects three statements.

Lemma 3. *(1) An optimal solution to Φ also maximizes $\Phi_{\bigcirc}$, i.e., $\Phi_{\bigcirc}(\mathcal{S}^*_{\bigcirc}) = \Phi_{\bigcirc}(\mathcal{S}^*)$.*

(2) It holds that $\Phi_{\bigcirc}(\mathcal{S}^*_{\bigcirc}) \geq l \cdot p \cdot \mathrm{DKSH}^*_d$, where DKSH^*_d is the optimal value of DENSEST-k-SUB-d-HYPERGRAPH in Q and $p = d!/d^d$.

(3) Given $\mathcal{S} \in \Sigma$, we can, in polynomial time, build a feasible solution S of Q such that $|E(S)| \geq \Phi_{\bigcirc}(\mathcal{S})/(\lambda l)$.

The first statement says that an optimal solution to Φ also maximizes $\Phi_{\bigcirc}$. This essentially holds since we have constructed the instance in a way that the contribution of circle-nodes dominates. The second statement says that there exists a feasible solution to P which achieves at least a multiple of $l \cdot p$ of the objective value in DENSEST-k-SUB-d-HYPERGRAPH with $p = d!/d^d$. In the third statement, we observe that from a feasible solution to P, we can construct a feasible solution to Q while loosing only a factor of λl in objective value. Using Lemma 3, we are now ready to prove Theorem 1.

Proof of Theorem 1. Let $Q = (G,k)$ be an instance of the DENSEST-k-SUB-d-HYPERGRAPH problem and let $P := (\overline{G} = (\overline{V}, \overline{E}), \mathcal{P}, \mathcal{I}, \overline{k}) = \tau(G,k)$ be the instance of the μ-ν-BALANCE problem obtained by the transform τ. For brevity, let $n := |V|$ and $\overline{n} := |\overline{V}|$. Moreover, let $\mathcal{S}$ be an $\alpha(|\overline{V}|)$-approximate solution to P, that is $\Phi(\mathcal{S}) \geq \alpha(|\overline{V}|)\Phi(\mathcal{S}^*)$. We show how to construct a $\beta(n)$-approximate solution S to Q.

Using Lemma 3, (3), we obtain a feasible solution S to Q with $|E(S)| \geq \Phi_{\bigcirc}(\mathcal{S})/(\lambda l)$. We proceed by lower-bounding $\Phi_{\bigcirc}(\mathcal{S})$. We can w.l.o.g. assume that $\mathcal{S} \cap V_{\bigcirc} = \emptyset$ and that $\Phi_{\bigcirc}(\mathcal{S}) \geq l$. Indeed, if $\Phi_{\bigcirc}(\mathcal{S}) < l$ then $\Phi_{\bigcirc}(\mathcal{S}) = 0$ and we can build in polynomial-time a better solution by identifying one edge $(v_1, \ldots, v_d)$ and propagating campaign i in v_i. This further implies that $\Phi_{\bigcirc}(\mathcal{S}) \geq \Phi_{\square}(\mathcal{S})$ as $l > n \geq \Phi_{\square}(\mathcal{S})$. We obtain

$$\begin{aligned}\Phi_{\bigcirc}(\mathcal{S}) &\geq \frac{\Phi(\mathcal{S})}{2} \geq \frac{\alpha(\overline{n})}{2} \cdot \Phi(\mathcal{S}^*) \geq \frac{\alpha(\overline{n})}{2} \cdot \Phi_{\bigcirc}(\mathcal{S}^*)\\ &= \frac{\alpha(\overline{n})}{2} \cdot \Phi_{\bigcirc}(\mathcal{S}^*_{\bigcirc}) \geq \frac{\alpha(\overline{n}) \cdot l \cdot p}{2} \cdot \mathrm{DKSH}^*_d,\end{aligned}$$

using Lemma 3, (1) and (2) in the last two steps. In summary, we have $|E(S)| \geq \frac{\alpha(\overline{n}) \cdot p}{2\lambda} \mathrm{DKSH}^*_d$. Note that $2\lambda/p$ is a constant. Moreover, by using that $2 \leq n \leq \overline{n} \leq 2\lambda n^{d+1}$ and $\lambda \leq \mu^d \leq n^d$, we get that $\overline{n} \leq n^{2d+2}$.

Case $d = 2$: Since g is non-increasing, it follows that

$$\alpha(\overline{n}) = \overline{n}^{-g(\overline{n})} \geq n^{-(2d+2)g(n)} = n^{-6g(n)},$$

completing this case.

Case $d \geq 3$: In this case

$$\alpha(\overline{n}) = \overline{n}^{-\epsilon(d)} \geq n^{-(2d+2)\epsilon(d)},$$

completing this case. □

4 Approximation Algorithm for the Heterogeneous Case

Our approach for maximizing $\Phi(\mathcal{S})$ is to decompose it as $\Phi(\mathcal{S}) = \Phi^0(\mathcal{S}) + \Phi^{\geq 1}(\mathcal{S})$ and work on each summand separately. In this section, we give two different algorithms GreedyTuple and GreedyIter for maximizing $\Phi^{\geq 1}(\mathcal{S})$. These two algorithms are inspired by a similar approach for the so-called maximum coverage with pairs problem (D'Angelo, Olsen, and Severini 2019).

The complete approximation algorithm then works as follows: Using GreedyTuple$(\epsilon, \delta/2, 1, \mathcal{I}, \nu, k)$, we obtain a set $\mathcal{S}^1$ that with probability $1 - \delta/2$ satisfies $\Phi^{\geq 1}(\mathcal{S}^1) \geq \alpha_1 \cdot \Phi^{\geq 1}(\mathcal{S}^*_{\geq 1})$, where $\mathcal{S}^*_{\geq 1}$ denotes an optimal solution of size k to maximizing $\Phi^{\geq 1}$ and α_1 is the approximation factor that will be achieved by GreedyTuple, see Theorem 3. On the other hand GreedyIter$(\epsilon, \delta/2, \mathcal{I}, \nu, k)$ outputs a set $\mathcal{S}^2$ that with probability $1 - \delta/2$ satisfies $\Phi(\mathcal{S}^2) \geq \alpha_2 \cdot \Phi^{\geq 1}(\mathcal{S}^*_{\geq 1})$, where α_2 is as in Theorem 4. Now, we define $\mathcal{S}'$ to be the solution that achieves the maximum $\max\{\Phi(\mathcal{S}^1), \Phi(\mathcal{S}^2)\}$ and $\mathcal{S}$ to be the solution that achieves the maximum $\max\{\Phi(\emptyset), \Phi(\mathcal{S}')\}$. It follows that

$$2\Phi(\mathcal{S}) \geq \Phi(\emptyset) + \Phi(\mathcal{S}') \geq \Phi^0(\emptyset) + \sqrt{\alpha_1 \cdot \alpha_2}\Phi^{\geq 1}(\mathcal{S}^*_{\geq 1}),$$

using that the maximum $\Phi(\mathcal{S}')$ is lower bounded by the geometric mean of $\Phi(\mathcal{S}^1)$ and $\Phi(\mathcal{S}^2)$, which are in turn lower bounded by $\alpha_1 \cdot \Phi^{\geq 1}(\mathcal{S}^*_{\geq 1})$ and $\alpha_2 \cdot \Phi^{\geq 1}(\mathcal{S}^*_{\geq 1})$, respectively. Now, let $\mathcal{S}^*$ be an optimal solution of size k to maximizing Φ. Using that the empty set maximizes Φ^0, we have $\Phi^0(\emptyset) \geq \Phi^0(\mathcal{S}^*)$. Furthermore, $\Phi^{\geq 1}(\mathcal{S}^*_{\geq 1}) \geq \Phi^{\geq 1}(\mathcal{S}^*)$, thus

$$\Phi(\mathcal{S}) \geq \frac{\sqrt{\alpha_1 \alpha_2}}{2}(\Phi^0(\mathcal{S}^*) + \Phi^{\geq 1}(\mathcal{S}^*)) = \frac{\sqrt{\alpha_1 \alpha_2}}{2}\Phi(\mathcal{S}^*).$$

Plugging in α_1 and α_2, see Theorems 3 and 4, and using $k^{\nu-2} \geq \binom{k-1}{\nu-2}$, we get the following theorem.

Theorem 2. *Let $0 < \epsilon < 1$ and $\delta \leq 1/2$. There is an algorithm that, with probability $1 - \delta$, outputs a solution $\mathcal{S}$ that satisfies*

$$\Phi(\mathcal{S}) \geq \Big(1 - \frac{1}{e} - \epsilon\Big)^{\frac{\nu}{2}} \Big(\frac{1}{2|V|}\Big)^{\frac{\nu-2}{2}} \nu^{-\frac{2\nu-3}{2}} \cdot \Phi(\mathcal{S}^*),$$

where $\mathcal{S}^*$ denotes a solution of size k maximizing $\Phi(\cdot)$.

We remark that this result is mostly interesting for the case where $\nu = 3$. Indeed, in this case we obtain an algorithm with an approximation ratio of order $n^{-1/2}$. It remains to describe algorithms GreedyTuple and GreedyIter.

Greedily Picking Tuples. This paragraph presents GreedyTuple (Algorithm 1). For given ℓ, the algorithm computes a solution maximizing $\Phi^{\geq \ell}$. For $\ell = \nu - 1$, it is identical to the standard greedy hill climbing algorithm. For the general case of $\ell \leq \nu - 1$, we show the following theorem.

Theorem 3. *Let $\epsilon \in (0, 1)$, $\delta \leq 1/2$, and $\ell \in [1, \nu - 1]$. If $k \geq 2\nu/\epsilon$, with probability at least $1 - \delta$, the algorithm* GreedyTuple$(\epsilon, \delta, \ell, \mathcal{I}, \nu, k)$ *returns a solution $\mathcal{S}$ with*

$$\Phi^{\geq \ell}(\mathcal{S}) \geq \frac{1 - \frac{1}{e} - \epsilon}{\binom{k-1}{\nu-\ell-1}} \cdot \Phi^{\geq \ell}(\mathcal{S}^*_{\geq \ell}),$$

where $\mathcal{S}^*_{\geq \ell}$ denotes a solution of size k maximizing $\Phi^{\geq \ell}$.

Algorithm 1: GreedyTuple$(\epsilon, \delta, \ell, \mathcal{I}, \nu, k)$

$t \leftarrow \lceil \frac{k}{\nu-\ell} \rceil \binom{|\hat{V}|}{\nu-\ell}$; $\delta' \leftarrow \frac{\delta}{t}$; $\epsilon' \leftarrow \frac{\epsilon}{2e \cdot \binom{k}{\nu-\ell}}$; $\mathcal{S} \leftarrow \emptyset$;
while $|\mathcal{S}| \leq k - (\nu - \ell)$ **do**
  Add $\arg\max_{\tau \subseteq \hat{V}, |\tau| = \nu-\ell}\{$ $\texttt{approx}(\Phi^{\geq \ell}, \mathcal{S} \cup \tau, \mathcal{I}, \nu, \epsilon', \delta')\}$ to $\mathcal{S}$;
return $\mathcal{S}$;

The main idea underlying GreedyTuple is very much related to the standard greedy algorithm. That is instead of greedily adding elements to the set $\mathcal{S}$, here we greedily add tuples of size $\nu - \ell$ using the key observation from Lemma 1 that there is always an element that yields a $1/\binom{k}{\nu-\ell}$ fraction of the optimum possible progress. Thus, (ignoring the approximation issue) every step of the algorithm incurs a factor of $(1 - (1 - 1/\binom{k}{\nu-\ell}))$ to the approximation ratio.

Being Iteratively Greedy. This paragraph presents GreedyIter (Algorithm 2). Recall that we defined the function $\Phi^{\geq \ell}$. We now extend this notation by letting

$$\Phi^{\geq \ell}_{\beta}(\mathcal{R}, \mathcal{S}) := \mathrm{E}_{\mathcal{X}}[\mathrm{NoSM}_{\mu,\beta}(\rho_{\mathcal{X}}(\mathcal{R} \cup \mathcal{S}) \setminus \textstyle\bigcup_{j=0}^{\ell-1} V^{j,\mathcal{R}}_{\mathcal{X}})]$$

where $\ell \in [\nu - 1]$ and $\beta \in [\nu]$; we will mainly be working with $\beta = \ell + 1$. This function measures the expected number

Algorithm 2: GREEDYITER$(\epsilon, \delta, \mathcal{I}, \nu, k)$

$\delta' \leftarrow \delta/\nu$; $\epsilon' \leftarrow \epsilon/2$; $\mathcal{R}^{[1]} \leftarrow \mathcal{I}$;
for $\ell = 1, \ldots, \nu - 1$ **do**
  $\mathcal{S}^{[\ell]} \leftarrow$
    GREEDY$(\Phi^{\geq \ell}_{\ell+1}(\mathcal{R}^{[\ell]}, \cdot), \epsilon', \delta', \mathcal{R}^{[\ell]}, \ell+1, \lfloor \frac{k}{\nu-1} \rfloor)$;
  $\mathcal{R}^{[\ell+1]} \leftarrow \mathcal{R}^{[\ell]} \cup \mathcal{S}^{[\ell]}$;
return $\bigcup_{i=1}^{\nu-1} \mathcal{S}^{[i]}$;

of nodes reached by at least β campaigns from $\mathcal{R} \cup \mathcal{S}$ among the nodes that have been reached by at least ℓ campaigns from $\mathcal{R}$.

We apply the following iterative scheme in GREEDYITER: For ℓ from 1 to $\nu - 1$, find sets $\mathcal{S}^{[\ell]}$ of size $\lfloor k/(\nu - 1) \rfloor$ maximizing $\Phi^{\geq \ell}_{\ell+1}(\mathcal{R}^{[\ell]}, \cdot)$, where $\mathcal{R}^{[\ell]} := \mathcal{I} \cup \bigcup_{j=1}^{\ell-1} \mathcal{R}^{[j]}$. That is, in the ℓ^{th} iteration, we maximize the number of nodes reached by $\ell+1$ campaigns that have previously been reached by at least ℓ campaigns. The approach is motivated by the observation that, for any $\ell \in [\nu - 1]$ and initial sets $\mathcal{R}$, the function $\Phi^{\geq \ell}_{\ell+1}(\mathcal{R}, \mathcal{S})$ is monotone and submodular in $\mathcal{S}$, compare with Section 2 where we used this fact for $\ell = \nu - 1$. Using Lemma 2 applied to $\Phi^{\geq \ell}_{\ell+1}(\mathcal{R}, \cdot)$ with $\nu = \ell + 1$ we get that the standard greedy algorithm can be used in order to obtain a $(1 - 1/e - \epsilon)$-approximate solution for $\Phi^{\geq \ell}_{\ell+1}(\mathcal{R}, \cdot)$. This leads to the following theorem regarding the approximation ratio achieved by GREEDYITER.

Theorem 4. *Let $0 < \epsilon < 1$ and $\delta \leq 1/2$. With probability $1 - \delta$, GREEDYITER$(\epsilon, \delta, \mathcal{I}, \nu, k)$ returns $\mathcal{S}$ satisfying*

$$\Phi(\mathcal{S}) \geq \frac{(1 - \frac{1}{e} - \epsilon)^{\nu-1}}{\nu^{2\nu-3}} \left(\frac{k}{2|V|}\right)^{\nu-2} \cdot \Phi^{\geq 1}(\mathcal{S}^*_{\geq 1}),$$

where $\mathcal{S}^*_{\geq 1}$ denotes a solution of size k maximizing $\Phi^{\geq 1}$.

5 Approximation Algorithm for the Correlated Case

We now turn to the correlated case. Recall that here the probability functions are identical for all campaigns, i.e., $p_1(e) = \ldots = p_\mu(e)$ for every edge $e \in E$. Moreover, the cascade processes are completely correlated, that is, for any edge (u, v), if node u propagates campaign i to v, then node u also propagates all other campaigns that reach it to v. In contrast to the heterogeneous case, in the correlated setting, we show that there is a constant factor approximation algorithm for μ-ν-BALANCE for any $\mu \geq \nu \geq 2$:

Theorem 5. *Let $0 < \epsilon < 1$ and $\delta \leq 1/2$. If $k \geq 2\nu/\epsilon$, there is an algorithm for the correlated setting that, with probability $1 - \delta$, outputs a solution $\mathcal{S}$ that satisfies $\Phi(\mathcal{S}) \geq (1 - 1/e - \epsilon)/(2(\nu + 1)) \cdot \Phi(\mathcal{S}^*)$, where $\mathcal{S}^*$ is an optimal solution of size k to maximizing Φ.*

Also here, the idea is to consider the decomposition of $\Phi(\mathcal{S})$ as $\Phi(\mathcal{S}) = \Phi^{\geq 1}(\mathcal{S}) + \Phi^0(\mathcal{S})$ for a solution $\mathcal{S}$. Clearly, $\Phi^0(\mathcal{S})$ is still optimal when $\mathcal{S} = \emptyset$. In order to approximate $\Phi^{\geq 1}$, we however apply a different technique: The idea is to pick ν campaigns and propagate them in the same $\lfloor k/\nu \rfloor$ nodes, exploiting that all campaigns spread in an identical manner. To that end, we consider one fictitious campaign, say campaign 0, spreading with the same probabilities as the others. We now maximize the number of reached nodes among all nodes that were (a) reached by at least one campaign from $\mathcal{I}$ and were (b) reached by no more than ν campaigns from $\mathcal{I}$. For this purpose, we had defined the function Ψ in Section 2. Recall that

$$\Psi(\mathcal{T}) := \mathrm{E}_{\mathcal{X}}[|(\rho^{(0)}_{X_0}(\mathcal{T}) \cap \bigcup_{j=1}^{\nu-1} V^j_{\mathcal{X}}) \cup \bigcup_{j=\nu}^{\mu} V^j_{\mathcal{X}}|]$$

and observe that $\Psi(\mathcal{T})$ measures the expected number of nodes that are either (1) reached by more than ν campaigns from $\mathcal{I}$ or (2) are reached by campaign 0 from $\mathcal{T}$ and were reached by at least one campaign from $\mathcal{I}$. Recall that Ψ is submodular and thus can be approximated to within $1 - 1/e - \epsilon$ for any $\epsilon > 0$ by the greedy algorithm. To prove Theorem 5 we make use of a combination of three claims: The first claim states that the optimum of Ψ has a higher value than the one of $\Phi^{\geq 1}$. The second statement of the lemma says that we loose a factor of roughly ν when choosing a set of size $\lfloor k/\nu \rfloor$ instead of k when maximizing Ψ (this is due to submodularity). The last statement shows that given a solution $\mathcal{T}$ of size $\lfloor k/\nu \rfloor$ for Ψ, one can build a certain solution $\mathcal{S}'$ of size k for $\Phi^{\geq 1}$ such that $\Psi(\mathcal{T}) = \Phi^{\geq 1}(\mathcal{S}')$.

6 Conclusion and Future Work

In this paper, we introduced the μ-ν-BALANCE problem which is a general form of the problem of balancing information exposure in a social network introduced by Garimella et al. (2017) for two campaigns. As in this original work of Garimella et al., we considered two different scenarios called the correlated and heterogeneous setting. While in the special case of $\mu = \nu = 2$ studied by Garimella et al., both settings lead to an identical picture from an approximation algorithms perspective, we showed that this situation changes drastically for $\mu \geq \nu > 2$: For the correlated setting, we designed a constant factor approximation algorithm for any values of μ and ν. In the heterogeneous case though, we obtained several approximation hardness results stating that it is unlikely to find an $n^{-g(n)}$-approximation algorithm with $g(n) = o(1)$ if $\nu \geq 3$ or even an $n^{-\epsilon}$-approximation algorithm where ϵ is a constant depending on ν if $\nu \geq 4$. We complemented this finding by designing an algorithm with approximation ratio $\Omega(n^{-1/2})$ for $\nu = 3$.

Several directions of future work are conceivable. First, it would be interesting to improve the approximation guarantee for the μ-ν-BALANCE problem in both settings, most importantly for the heterogeneous case with $\nu > 3$. Second, since the ν parameter in the problem is of a threshold flavor, it would be worth investigating a smoother objective function by considering various ν values, such that a node reached by ν_1 campaigns contributes more to the objective function than a node reached by $\nu_2 < \nu_1$ campaigns.

7 Acknowledgments

This work has been partially supported by the Italian MIUR PRIN 2017 Project ALGADIMAR "Algorithms, Games, and Digital Markets".

References

Applebaum, B. 2013. Pseudorandom generators with long stretch and low locality from random local one-way functions. *SIAM Journal on Computing* 42(5):2008–2037.

Aslay, Ç.; Matakos, A.; Galbrun, E.; and Gionis, A. 2018. Maximizing the diversity of exposure in a social network. In *Proceedings of the 2018 IEEE International Conference on Data Mining, ICDM 2018, Singapore, November 17-20, 2018*, 863–868.

Becker, R.; Corò, F.; D'Angelo, G.; and Gilbert, H. 2019. Balancing spreads of influence in a social network. *CoRR* abs/1906.00074.

Borgs, C.; Brautbar, M.; Chayes, J. T.; and Lucier, B. 2014. Maximizing social influence in nearly optimal time. In *Proceedings of the 25th Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2014, Portland, Oregon, USA, January 5-7, 2014*, 946–957.

Borodin, A.; Braverman, M.; Lucier, B.; and Oren, J. 2017. Strategyproof mechanisms for competitive influence in networks. *Algorithmica* 78(2):425–452.

Budak, C.; Agrawal, D.; and El Abbadi, A. 2011. Limiting the spread of misinformation in social networks. In *Proceedings of the 20th International Conference on World Wide Web, WWW 2011, Hyderabad, India, March 28 - April 1, 2011*, 665–674.

Chlamtác, E.; Dinitz, M.; Konrad, C.; Kortsarz, G.; and Rabanca, G. 2018. The densest k-subhypergraph problem. *SIAM Journal on Discrete Mathematics* 32(2):1458–1477.

Conover, M. D.; Ratkiewicz, J.; Francisco, M. R.; Gonçalves, B.; Menczer, F.; and Flammini, A. 2011. Political polarization on Twitter. In *Proceedings of the 5th International Conference on Weblogs and Social Media, ICWSM 2011, Barcelona, Catalonia, Spain, July 17-21, 2011*.

D'Angelo, G.; Olsen, M.; and Severini, L. 2019. Coverage centrality maximization in undirected networks. In *Proceedings of the 33rd AAAI Conference on Artificial Intelligence, AAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019*, 501–508.

Del Vicario, M.; Bessi, A.; Zollo, F.; Petroni, F.; Scala, A.; Caldarelli, G.; Stanley, H. E.; and Quattrociocchi, W. 2016. The spreading of misinformation online. *Proceedings of the National Academy of Sciences* 113(3):554–559.

Fish, B.; Bashardoust, A.; Boyd, D.; Friedler, S. A.; Scheidegger, C.; and Venkatasubramanian, S. 2019. Gaps in information access in social networks? In *Proceedings of the 2019 World Wide Web Conference, WWW 2019, San Francisco, CA, USA, May 13-17, 2019*, 480–490.

Garimella, K.; Gionis, A.; Parotsidis, N.; and Tatti, N. 2017. Balancing information exposure in social networks. In *Advances in Neural Information Processing Systems, 30th Annual Conference on Neural Information Processing Systems, NIPS 2017, 4-9 December 2017, Long Beach, CA, USA*, 4666–4674.

Garimella, K.; Morales, G. D. F.; Gionis, A.; and Mathioudakis, M. 2018. Political discourse on social media: Echo chambers, gatekeepers, and the price of bipartisanship. In *Proceedings of the 2018 World Wide Web Conference, WWW 2018, Lyon, France, April 23-27, 2018*, 913–922.

Hochbaum, D. S., ed. 1997. *Approximation Algorithms for NP-hard Problems*. Boston, MA, USA: PWS Publishing Co.

Kempe, D.; Kleinberg, J. M.; and Tardos, É. 2003. Maximizing the spread of influence through a social network. In *Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2003, Washington, DC, USA, August 24 - 27, 2003*, 137–146.

Kempe, D.; Kleinberg, J. M.; and Tardos, É. 2005. Influential nodes in a diffusion model for social networks. In *Proceedings of the 32nd International Colloquium on Automata, Languages and Programming, ICALP 2005, Lisbon, Portugal, July 11-15, 2005*, 1127–1138.

Kempe, D.; Kleinberg, J. M.; and Tardos, É. 2015. Maximizing the spread of influence through a social network. *Theory of Computing* 11:105–147.

Manurangsi, P. 2017. Almost-polynomial ratio ETH-hardness of approximating densest k-subgraph. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2017, Montreal, QC, Canada, June 19-23, 2017*, 954–961.

Matakos, A., and Gionis, A. 2018. Tell me something my friends do not know: Diversity maximization in social networks. In *Proceedings of the 2018 IEEE International Conference on Data Mining, ICDM 2018, Singapore, November 17-20, 2018*, 327–336.

Raghavendra, P., and Steurer, D. 2010. Graph expansion and the unique games conjecture. In *Proceedings of the 42nd ACM Symposium on Theory of Computing, STOC 2010, Cambridge, Massachusetts, USA, 5-8 June 2010*, 755–764.

Tang, F.; Liu, Q.; Zhu, H.; Chen, E.; and Zhu, F. 2014. Diversified social influence maximization. In *Proceedings of the 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2014, Beijing, China, August 17-20, 2014*, 455–459.

Thompson, A. 2016. Journalists and Trump voters live in separate online bubbles, MIT analysis shows. https://www.vice.com/en_us/article/d3xamx/journalists-and-trump-voters-live-in-separate-online-bubbles-mit-analysis-shows. Accessed: 2019-09-03.

Tsang, A.; Wilder, B.; Rice, E.; Tambe, M.; and Zick, Y. 2019. Group-fairness in influence maximization. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019*, 5997–6005.