The Thirty-Second AAAI Conference
on Artificial Intelligence (AAAI-18)

Maximizing Activity in Ising Networks via the TAP Approximation

Christopher W. Lynn
Department of Physics & Astronomy
University of Pennsylvania
Philadelphia, PA 19104, USA

Daniel D. Lee
Department of Electrical & Systems Engineering
University of Pennsylvania
Philadelphia, PA 19104, USA

Abstract

A wide array of complex biological, social, and physical systems have recently been shown to be quantitatively described by Ising models, which lie at the intersection of statistical physics and machine learning. Here, we study the fundamental question of how to optimize the state of a networked Ising system given a budget of external influence. In the continuous setting where one can tune the influence applied to each node, we propose a series of approximate gradient ascent algorithms based on the Plefka expansion, which generalizes the naïve mean field and TAP approximations. In the discrete setting where one chooses a small set of influential nodes, the problem is equivalent to the famous influence maximization problem in social networks with an additional stochastic noise term. In this case, we provide sufficient conditions for when the objective is submodular, allowing a greedy algorithm to achieve an approximation ratio of $1-1/e$. Additionally, we compare the Ising-based algorithms with traditional influence maximization algorithms, demonstrating the practical importance of accurately modeling stochastic fluctuations in the system.

Introduction

The last 10 years have witnessed a dramatic increase in the use of maximum entropy models to describe a diverse range of real-world systems, including networks of neurons in the brain (Schneidman et al. 2006; Ganmor, Segev, and Schneidman 2011), flocks of birds in flight (Bialek et al. 2012), and humans interacting in social networks (Lynn et al. 2017; Galam 2008), among an array of other social and biological applications (Kapur 1989; Phillips, Anderson, and Schapire 2006; Mora et al. 2010; Lezon et al. 2006). Broadly speaking, the maximum entropy principle allows scientists to formalize the hypothesis that large-scale patterns in complex systems emerge organically from an aggregation of simple fine-scale interactions between individual elements (Jaynes 1957). Indeed, intelligence itself, either naturally-occurring in the human brain and groups of animals (Hillis 1988) or artificially constructed in learning algorithms and autonomous systems (Mataric 1993; Namatame, Kurihara, and Nakashima 2007), is increasingly viewed as an emergent phenomenon (Lévy 1997), the result of repeated underlying interactions between populations of smaller elements.

Given the wealth of real-world systems that are quantitatively described by maximum entropy models, it is of fundamental practical and scientific interest to understand how external influence affects the dynamics of these systems. Fortunately, all maximum entropy models are similar, if not formally equivalent, to the Ising model, which has roots in statistical physics (Brush 1967) and has a rich history in machine learning as a model of neural networks (Coughlin and Baran 1995). The state of an Ising system is described by the average activity of its nodes. For populations of neurons in the brain, an active node represents a spiking neuron while inactivity represents silence. In the context of humans in social networks, node activity could represent the sending of an email or the consumption of online entertainment, while inactivity represents moments in which an individual does not perform an action. By applying external influence to a particular node, one can shift the average activity of that node. Furthermore, this targeted influence also has indirect effects on the rest of the system, mediated by the underlying network of interactions. For example, if an individual is incentivized to send more emails, this shift in behavior induces responses from her neighbors in the social network, resulting in increased activity in the population as a whole.

As a first step toward understanding how to control such complex systems, we study the problem of maximizing the total activity of an Ising network given a budget of external influence. This so-called *Ising influence maximization* problem was originally proposed in the context of social networks (Lynn and Lee 2016), where it has a clear practical interpretation: If a telephone company or an online service wants to maximize user activity, how should it distribute its limited marketing resources among its customers? However, we emphasize that the broader goal—to develop a unifying control theory for understanding the effects of external influence in complex systems—could prove to have other important applications, from guiding healthy brain development (Goddard, McIntyre, and Leech 1969) and intervening to alleviate diseased brain states (Goddard 1967) to anticipating trends in financial markets (Mantegna and Stanley 1999) and preventing viral epidemics (Pastor-Satorras and Vespignani 2001).

We divide our investigation into two settings: (i) the continuous setting where one can tune the influence applied to each node, and (ii) the discrete setting in which one forces activation upon a small set of influential nodes. In the continuous setting, we propose a gradient ascent algorithm and give novel conditions for when the objective is concave. We then present a series of increasingly-accurate approximation algorithms based on an advanced approximation technique known as the Plefka expansion (Plefka 1982). The Plefka expansion generalizes the naïve mean field and TAP approximations, and, in theory, can be extended to arbitrary order (Yedidia 2001).

In the discrete setting, it was recently shown that Ising influence maximization is closely related to the famous influence maximization problem in social networks (Kempe, Kleinberg, and Tardos 2003) with the addition of a natural stochastic noise term (Lynn and Lee 2017). Here, we provide novel conditions for when the total activity of the system is submodular with respect to activated nodes. This result guarantees that a greedy algorithm achieves a $1 - 1/e$ approximation to the optimal choice of nodes. We compare our greedy algorithm with traditional influence maximization techniques, demonstrating the importance of accurately accounting for stochastic noise.

Related Work

Ising influence maximization was originally proposed in the context of human activity in social networks (Lynn and Lee 2016). However, a recent surge in the use of Ising models to describe other biological and physical systems significantly expands the problem's applicability (Stein, Marks, and Sander 2015).

Ising influence maximization was originally studied in the continuous setting under the naïve mean field approximation. Since the Plefka expansion generalizes the mean field approximation to increasing levels of accuracy, our work represents a principled improvement over existing techniques.

In the discrete setting, it was recently shown that Ising influence maximization is closely related to standard influence maximization (Lynn and Lee 2017), which was first studied in the context of viral marketing (Domingos and Richardson 2001). Kempe et al. (Kempe, Kleinberg, and Tardos 2003) proposed influence maximization as a discrete optimization problem and presented a greedy algorithm with approximation guarantees. Significant subsequent research has focused on developing efficient greedy and heuristic techniques (Leskovec et al. 2007; Chen, Wang, and Yang 2009; Chen, Wang, and Wang 2010). Here, we do not claim to provide improvements over these algorithms in the context of standard influence maximization. Instead, we focus on developing analogous techniques that are suitable for the Ising model.

Ising Influence Maximization

In the study of complex systems, if we look through a sufficiently small window in time, the actions of individual elements appear binary—either human i sent an email ($\sigma_i = 1$) or she did not ($\sigma_i = -1$). The binary vector $\boldsymbol{\sigma} = \{\sigma_i\} \in \{\pm 1\}^n$ represents the activity of the entire system at a given moment in time, where n is the size of the system.

Many complex systems in the biological and social sciences have recently been shown to be quantitatively described by the Ising model from statistical physics. The Ising model is defined by the Boltzmann distribution over activity vectors:

$$P(\boldsymbol{\sigma}) = \frac{1}{Z} \exp\left(\frac{1}{2}\sum_{i \neq j} J_{ij}\sigma_i\sigma_j + \sum_i b_i\sigma_i\right), \qquad (1)$$

where Z is a normalization constant. The parameters $J = \{J_{ij}\}$ define the network of interactions between elements and the parameters $\boldsymbol{b} = \{b_i\}$ represent individual biases, which can be altered by application of targeted external influence. For example, if J defines the network of interactions in a population of email users, then $\boldsymbol{b}$ represents the intrinsic tendencies of users to send emails, which can be shifted by incentivizing or disincentivizing email use.

Problem Statement

The total average activity of a network with bias $\boldsymbol{b}$ is denoted $M(\boldsymbol{b}) = \sum_i \langle \sigma_i \rangle$, where $\langle \cdot \rangle$ denotes an average over the Boltzmann distribution (1). In what follows, we assume that the interactions J and initial bias $\boldsymbol{b}^0$ are known. We note that this assumption is not restrictive since there exist an array of advanced techniques in machine learning (Ackley, Hinton, and Sejnowski 1985) and statistical mechanics (Aurell and Ekeberg 2012) for learning Ising parameters directly from observations of a system.

We study the problem of maximizing the total activity M with respect to an additional external influence $\boldsymbol{h}$, subject to the budget constraint $|\boldsymbol{h}|_p = \left(\sum_i |h_i|^p\right)^{1/p} \leq H$, where H is the budget of external influence.

Problem 1 (Ising influence maximization). Given an Ising system defined by J and $\boldsymbol{b}^0$, and a budget H, find an optimal external influence $\boldsymbol{h}^*$ satisfying

$$\boldsymbol{h}^* = \underset{|\boldsymbol{h}|_p \leq H}{\arg\max}\, M(\boldsymbol{b}^0 + \boldsymbol{h}). \qquad (2)$$

We point out that the norm p plays an important role. If $p = 1, 2, 3, \ldots$, then one is allowed to tune the influence on each node continuously. On the other hand, if $p = 0$, then $|\boldsymbol{h}|_0$ counts the number of non-zero elements in $\boldsymbol{h}$. In this case, one chooses a subset of $\lfloor H \rfloor$ nodes $\{i\}$ to activate with probability one by sending $\{h_i\} \to \infty$.

The Plefka Expansion

Since the Ising model has remained unsolved for all but a select number of special cases, tremendous interdisciplinary effort has focused on developing tractable approximation techniques. Here, we present an advanced approximation method known as the Plefka expansion (Yedidia 2001). The Plefka expansion is not an approximation itself, but is rather a principled method for deriving a series of increasingly accurate approximations, the first two orders of which are the naïve mean-field (MF) and Thouless-Anderson-Palmer (TAP) approximations. In subsequent sections, we will use the Plefka expansion to approximately solve the Ising influence maximization problem in (2).

Calculations in the Ising model, such as the average activity $\langle \sigma_i \rangle$, generally require summing over all 2^n binary activity vectors. To get around this exponential dependence on system size, the Plefka expansion provides a series of approximations based on the limit of weak interactions $|J_{ij}| \ll 1$. Each order α of the expansion generates a set of self-consistency equations $m_i = f_i^{(\alpha)}(\boldsymbol{m})$, where m_i approximates the average activity $\langle \sigma_i \rangle$. Thus, for any order α of the Plefka expansion, the intractable problem of computing the averages $\langle \sigma_i \rangle$ is replaced by the manageable task of computing solutions to the corresponding self-consistency equations $\boldsymbol{m} = \boldsymbol{f}^{(\alpha)}(\boldsymbol{m})$. We point the interested reader to Appendix A for a detailed derivation of the Plefka expansion.

For a system with interactions J and bias $\boldsymbol{b}$, the first order in the expansion yields the naïve mean field approximation, summarized by the self-consistency equations

$$m_i = \tanh\left[b_i + \sum_j J_{ij} m_j\right] \triangleq f_i^{\text{MF}}(\boldsymbol{m}). \tag{3}$$

The second order in the Plefka expansion yields the TAP approximation,

$$m_i = \tanh\left[b_i + \sum_j J_{ij} m_j - m_i \sum_j J_{ij}^2 (1 - m_j^2)\right] \triangleq f_i^{\text{TAP}}(\boldsymbol{m}). \tag{4}$$

Higher-order approximations can be achieved by systematically including higher orders of J in the argument of $\tanh[\cdot]$. In Appendix B, we present a derivation of the third-order approximation, denoted TAP3.

The standard approach for computing solutions to the self-consistency equations $\boldsymbol{m} = \boldsymbol{f}^{(\alpha)}(\boldsymbol{m})$ is to iteratively apply $\boldsymbol{f}^{(\alpha)}$ until convergence is reached:

$$\boldsymbol{m} \leftarrow (1 - \gamma)\boldsymbol{m} + \gamma \boldsymbol{f}^{(\alpha)}(\boldsymbol{m}), \tag{5}$$

where $\gamma \in [0, 1]$ is the step size. The convergence of this procedure was rigorously examined in (Bolthausen 2014). In practice, we find that $\gamma \sim 0.01$ yields rapid convergence for most systems up to the third-order approximation.

The Continuous Setting

In this section, we study Ising influence maximization under a budget constraint $|\boldsymbol{h}|_p \leq H$, where $p = 1, 2, 3, \ldots$, yielding a continuous optimization problem where one can tune the external influence on each element in the system. We first present an exact gradient ascent algorithm and comment on its theoretical guarantees. We then demonstrate how the Plefka expansion can be used to approximate the gradient, yielding a series of increasingly accurate approximation algorithms.

Projected Gradient Ascent

We aim to maximize the total activity $M(\boldsymbol{b}^0 + \boldsymbol{h}) = \sum_i \langle \sigma_i \rangle$ with respect to the external influence $\boldsymbol{h}$. Thus, a crucially important concept is the response function

$$\chi_{ij} = \frac{\partial \langle \sigma_i \rangle}{\partial h_j} = \langle \sigma_i \sigma_j \rangle - \langle \sigma_i \rangle \langle \sigma_j \rangle, \tag{6}$$

Algorithm 1: Projected Gradient Ascent (PGA)

Input: Ising system defined by J and $\boldsymbol{b}^0$, budget H, norm p, and error ϵ;
Output: External influence $\boldsymbol{h}$;
Initialize: Choose $|\boldsymbol{h}^{(0)}|_p \leq H$, $k \leftarrow 0$;
while $\left|M(\boldsymbol{b}^0 + \boldsymbol{h}^{(k)}) - M(\boldsymbol{b}^0 + \boldsymbol{h}^{(k-1)})\right| > \epsilon$ **do**
  Choose step size η_k; $\boldsymbol{h}^{(k+1)} \leftarrow \pi_{|\boldsymbol{h}|_p \leq H}\left[\boldsymbol{h}^{(k)} + \eta_k \chi(\boldsymbol{b}^0 + \boldsymbol{h}^{(k)})^T \mathbf{1}\right]$;
  k++;
end
$\boldsymbol{h} \leftarrow \boldsymbol{h}^{(k)}$;

which quantifies the change in the activity of node i due to a shift in the influence on node j. The gradient of M with respect to $\boldsymbol{h}$ can be succinctly written $\nabla_{\boldsymbol{h}} M = \chi^T \mathbf{1}$, where $\mathbf{1}$ is the n-vector of ones.

In Algorithm 1 we present a projected gradient ascent algorithm PGA. Starting at a feasible choice for the external influence $\boldsymbol{h}^{(0)}$, PGA steps along the gradient $\chi^T \mathbf{1}$ and then projects back down to the space of feasible solutions $|\boldsymbol{h}|_p \leq H$. We note that for $p = 1, 2, 3, \ldots$, the space of feasible solutions is convex, and hence the projection $\pi_{|\boldsymbol{h}|_p \leq H}$ is well-defined and can be performed efficiently (Duchi et al. 2008).

Conditions for Optimality

The algorithm PGA efficiently converges to an ϵ-approximation of a local maximum of M in $O(1/\epsilon)$ iterations (Nesterov 2013). However, this local maximum could be arbitrarily far from the globally optimal solution. Here, we present a novel sufficient condition for when PGA is guaranteed to converge to a global maximum of M, subject to the proof of a long-standing conjecture.

Conjecture 2 (Sylvester 1983). Given an Ising system with non-negative interactions $J \geq 0$ and non-negative biases $\boldsymbol{b} \geq 0$, the average activity of each node $\langle \sigma_i \rangle$ is a concave function of the biases $\boldsymbol{b}$.

Theorem 3. If Conjecture 2 holds, then for any Ising system with non-negative interactions $J \geq 0$ and non-negative initial biases $\boldsymbol{b}^0 \geq 0$, PGA converges to a global maximum of the total activity M.

Proof. For Ising systems with positive couplings $J \geq 0$, the response function is non-negative $\{\chi_{ij}\} \geq 0$ (Griffiths 1967). This implies two things: (i) at least one global maximum of $M(\boldsymbol{b})$ occurs in the non-negative orthant of $\boldsymbol{b}$, and (ii) if $\boldsymbol{b}^0 \geq 0$, then $\boldsymbol{b}^0 + \boldsymbol{h}^{(k)}$ will be non-negative at every iteration k of PGA. If Conjecture 2 holds, then every local maximum in the non-negative orthant is a global maximum. Thus, PGA converges to a global maximum. □

We remark that Sylvester (Sylvester 1983) provides extensive experimental justification for Conjecture 2, and even

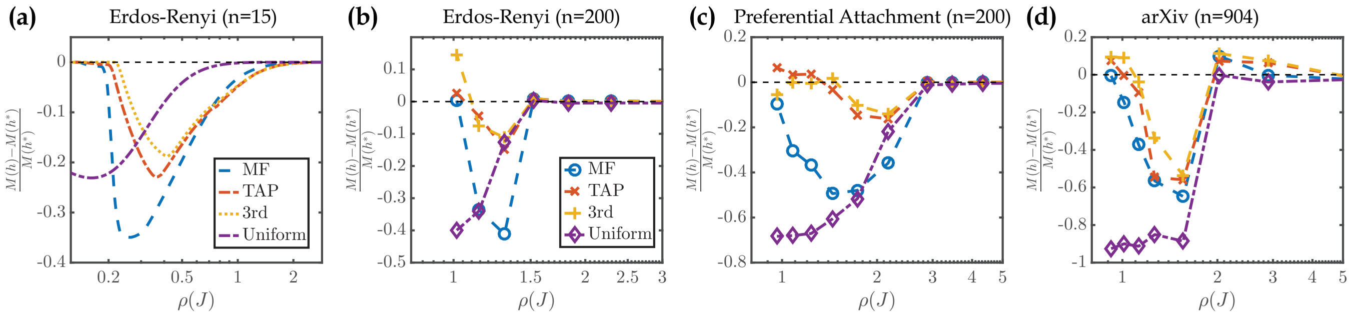

Figure 1: Performance of PGA for various orders of the Plefka expansion. (a) An Erdös-Rényi network with $n = 15$ nodes and budget $H = 1$. The total activity is calculated exactly using the Boltzmann distribution. (b) An Erdös-Rényi network with $n = 200$ nodes and budget $H = 10$. (c) A preferential attachment network with $n = 200$ nodes and budget $H = 10$. (d) A collaboration network of $n = 904$ physicists on the arXiv and budget $H = 20$. The total activities in (b-d) are estimated using Monte Carlo simulations. The benchmarks are PGA with the exact gradient for (a) and the gradient estimated using Monte Carlo simulations in (b-d).

proves Conjecture 2 in a number of limited cases. Additionally, we manually verified the veracity of Conjecture 2 in each of the experiments presented below. We also note that the sufficient conditions are plausible for many real-world scenarios. Positive interactions $J \geq 0$ imply that an action from one node will tend to induce an action from another node, a phenomenon that has been experimentally verified in small neuronal (Schneidman et al. 2006) and social (Lynn et al. 2017) networks. The more stringent condition is that $\boldsymbol{b}^0 \geq$, implying that each element in the network prefers activity over inactivity.

Approximating the Gradient via the Plefka Expansion

Since PGA requires calculating the response function χ at each iteration, an exact implementation scales exponentially with the size of the system. Here we show that the Plefka expansion can be used to approximate χ, yielding a series of efficient and increasingly-accurate gradient ascent algorithms.

Given a self-consistent approximation of the form $\boldsymbol{m} = \boldsymbol{f}^{(\alpha)}(\boldsymbol{m})$, where α denotes the order of the Plefka approximation, the response function is approximated by

$$\tilde{\chi}_{ij}^{(\alpha)} = \frac{\partial f_i^{(\alpha)}}{\partial h_j} + \sum_k \frac{\partial f_i^{(\alpha)}}{\partial m_k} \tilde{\chi}_{kj}^{(\alpha)}. \tag{7}$$

For all orders α of the Plefka expansion, we point out that $\partial f_i^{(\alpha)}/\partial h_j = (1 - m_i^2)\delta_{ij}$. Thus, defining $A_{ij} \triangleq (1 - m_i^2)\delta_{ij}$, and denoting the Jacobian of $\boldsymbol{f}^{(\alpha)}$ by $Df_{ij}^{(\alpha)} \triangleq \partial f_i^{(\alpha)}/\partial m_j$, the response function takes the particularly simple form

$$\tilde{\chi}^{(\alpha)} = (I - Df^{(\alpha)})^{-1} A, \tag{8}$$

where I is the identity matrix.

Thus, to approximate the gradient $\nabla_{\boldsymbol{h}} M \approx \tilde{\chi}^T \boldsymbol{1}$, one simply needs to calculate the Jacobian of $\boldsymbol{f}^{(\alpha)}$. Under the mean field approximation, the Jacobian takes the form $Df^{\text{MF}} = AJ$; and under the TAP approximation, we have

$$Df_{ij}^{\text{TAP}} = (1-m_i^2)\Big[J_{ij} + 2J_{ij}m_i m_j - \delta_{ij}\sum_k J_{ik}(1-m_k^2)\Big]. \tag{9}$$

We point the reader to Appendix B for a derivation of the third-order Jacobian.

Experimental Evaluation

In Fig. 1, we compare various orders of the Plefka approximation across a range of networks for the norm $p = 1$. We also compare with the uniform influence $\boldsymbol{h} = H/n\boldsymbol{1}$ as a baseline. In Fig. 1(a), the network is small enough that we can calculate the exact optimal solution $\boldsymbol{h}^*$, while for Figs. 1(b-d), we approximate $\boldsymbol{h}^*$ by running costly Monte Carlo simulations to estimate the gradient at each iteration of PGA. Similarly, we calculate the total activity M exactly using the Boltzmann distribution in Fig. 1(a), while in Figs. 1(b-d), we estimate M using Monte Carlo simulations.

For each network, we assume that the interactions are symmetric $J = J^T$ with uniform weights and that the initial bias is zero $\boldsymbol{b}^0 = 0$. We then study the performance of the various algorithms across a range of interaction strengths, summarized by the spectral radius $\rho(J)$. For $\rho(J) \ll 1$, the network is dominated by randomness and all influence strategies have little affect on the total activity. On the other hand, for $\rho(J) \gg 1$, the elements interact strongly and any positive influence induces the entire network to become active. Thus, the interesting regime is the "critical" region near $\rho(J) \approx 1$.

The striking success of the Plefka expansion is summarized by the fact that TAP and TAP3 consistently provide dramatic improvements over the naïve mean field algorithm studied in (Lynn and Lee 2016). Indeed, TAP and TAP3 consistently perform within 20% of optimal (except for the arXiv network) and sometimes even outperform the Monte Carlo algorithm benchmark in Figs. 1(b-d). Furthermore, while the Monte Carlo algorithm takes $\sim$ 10 minutes to

Algorithm 2: Greedy algorithm for choosing top H influential nodes in an Ising network (GI)

Input: Ising system defined by J and $\boldsymbol{b}^0$, budget H;
Output: Set of H influential nodes V;
Initialize: $V^{(1)} \leftarrow \{\}$;
for $k = 1, \dots, H$ **do**
  for *all nodes* $i \in \{1, \dots, n\}/V^{(k)}$ **do**
    Calculate total activity $\mathcal{M}(V^{(k)} \cup \{i\})$;
  end
  Choose node $i^* = \arg\max_i \mathcal{M}(V^{(k)} \cup \{i\})$;
  Add i^* to influential set $V^{(k+1)} \leftarrow V^{(k)} \cup \{i^*\}$;
end
$V \leftarrow V^{(k+1)}$;

complete in a network of size 200, PGA with the TAP and TAP3 approximations converges within ~ 5 seconds.

Discrete Setting

We now consider the discrete setting corresponding to a budget constraint of the form $|\boldsymbol{h}|_0 \leq H$. In this setting, one is allowed to apply infinite external influence to a set of $\lfloor H \rfloor$ nodes in the system, activating them with probability one; that is, one chooses a set of nodes $V = \{i\}$ for which we impose $\langle \sigma_i \rangle = 1$ by taking $h_i \to +\infty$. We begin by presenting a greedy algorithm that selects the single node at each iteration that yields the largest increase in the total activity M. We then provide novel conditions for when M is submodular in the selected nodes, which guarantees that our greedy algorithm is within $1 - 1/e$ of optimal. Finally, we comment on the relationship between (discrete) Ising influence maximization and traditional influence maximization in viral marketing, and we present experiments comparing our greedy algorithm with traditional techniques.

A Greedy Algorithm

We aim to maximize the total activity M with respect to a set V of activated nodes of size $|V| = H$ (assuming H is integer). To eliminate confusion, we denote by $\mathcal{M}(V)$ the total activity of the system after activating the nodes in V, assuming that the couplings J and initial bias $\boldsymbol{b}^0$ are already known.

Since there are $\binom{n}{H} \sim n^H$ possible choices for V, an exhaustive search for the optimal set is generally infeasible. On the other hand, we can simplify our search by looking at one node at a time and iteratively adding to V the single node that increases $\mathcal{M}$ the most. This approximate greedy approach was made famous in traditional influence maximization in the context of viral marketing by Kempe et al. (Kempe, Kleinberg, and Tardos 2003). In Algorithm 2 we propose an analogous algorithm for computing the top H influential nodes in an Ising system.

Theoretical Guarantee

The greedy algorithm GI efficiently chooses an approximate set V of influential nodes in $O(nH)$ iterations. However, V could be arbitrarily far from the globally optimal set of nodes. Here, we present novel conditions for when $\mathcal{M}$ is monotonic and submodular in V, and, hence, GI is guaranteed to compute a $1 - 1/e$ approximation of the optimal set of nodes. The proof is based on the famous GHS inequality from statistical physics.

Theorem 4 (Griffiths, Hurst, and Sherman 1970). Given an Ising system with non-negative interactions $J \geq 0$ and non-negative biases $\boldsymbol{b} \geq 0$, we have $\frac{\partial^2 \langle \sigma_i \rangle}{\partial b_j \partial b_k} \leq 0$ for all elements i, j, k.

We note that the GHS inequality guarantees a limited type of concavity of $\langle \sigma_i \rangle$ in the direction of positive bias $\boldsymbol{b}$. While this was not enough to prove that PGA is optimal in the continuous setting, it is strong enough to guarantee that $\mathcal{M}$ is submodular in the discrete setting.

Theorem 5. For an Ising system with non-negative interactions $J \geq 0$ and non-negative initial biases $\boldsymbol{b}^0 \geq 0$, the total activity $\mathcal{M}$ is monotonic and submodular in the activated nodes V.

Proof. Monotonicity is guaranteed for any system with non-negative interactions in (Griffiths 1967). To prove submodularity, we first introduce the notation $h_i^V \in \{0, 1\}$ if i is or is not in V. Then we note that $\mathcal{M}(V) \equiv \lim_{c \to \infty} M(\boldsymbol{b}^0 + c\boldsymbol{h}^V)$. Since M is non-negative and concave in the direction of positive bias for $J \geq 0$ and $\boldsymbol{b}^0 \geq 0$, M it is subadditive. Thus, for any set V of nodes and any two nodes $i, j \notin V$, we have

$$\begin{aligned} &M(\boldsymbol{b}^0 + c(\boldsymbol{h}^V + \boldsymbol{h}^{\{i\}})) + M(\boldsymbol{b}^0 + c(\boldsymbol{h}^V + \boldsymbol{h}^{\{j\}})) \\ &\geq M(\boldsymbol{b}^0 + c(\boldsymbol{h}^V + \boldsymbol{h}^{\{i\}} + \boldsymbol{h}^{\{j\}})) + M(\boldsymbol{b}^0 + c\boldsymbol{h}^V). \end{aligned} \tag{10}$$

Taking $c \to \infty$, we find that

$$\mathcal{M}(V \cup \{i\}) + \mathcal{M}(V \cup \{j\}) \geq \mathcal{M}(V \cup \{i, j\}) + \mathcal{M}(V), \tag{11}$$

which is the formal definition for submodularity. □

Relationship Between the Linear Threshold and Ising Models

It was recently established that the discrete version of the Ising influence maximization problem is closely related to traditional influence maximization in social networks (Lynn and Lee 2017). In traditional influence maximization, one aims to maximize the spread of activations under a viral model, such as linear threshold (LT) or independent cascade. For example, the LT model is defined by the deterministic dynamics

$$\sigma_i^{(t+1)} \leftarrow \text{sign}\Big[\sum_j J_{ij}\sigma_j^{(t)} + b_i\Big]. \tag{12}$$

Typically, one considers activation variables $\sigma_i' \in \{0, 1\}$ instead of $\sigma_i = \pm 1$, which can be accomplished by a simple change of parameters $J_{ij}' \leftarrow 2J_{ij}$ and $b_i' \leftarrow b_i - \sum_j J_{ij}$.

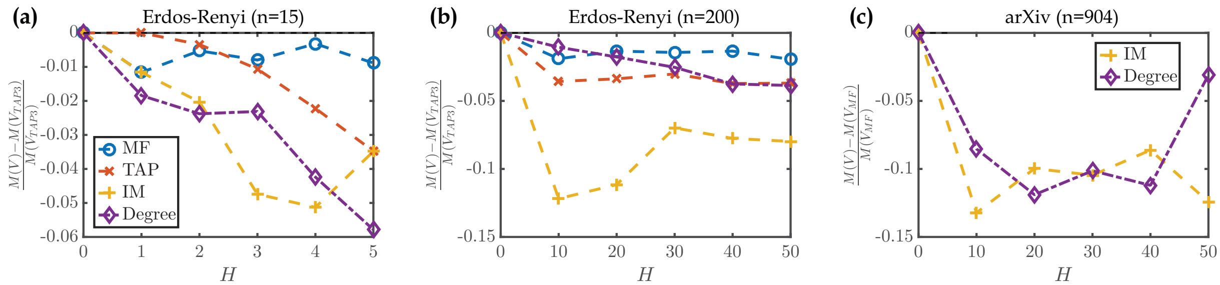

Figure 2: Comparison of the total Ising activity for greedy algorithms using various orders of the Plefka expansion. For each network, we ensure $\sum_j J_{ij} \leq 1/2$ and we average over many draws of the initial bias $\{b_i^0\} \sim \mathcal{U}[-1/2, 1/2]$. (a) An Erdös-Rényi network with $n = 15$ nodes. The total activity is calculated exactly using the Boltzmann distribution. (b) An Erdös-Rényi network with $n = 200$ nodes. (c) A collaboration network of $n = 904$ physicists on the arXiv. The total activities in (b-c) are estimated using Monte Carlo simulations. In (a-b) the benchmark is TAP3, while for (c) the benchmark is MF.

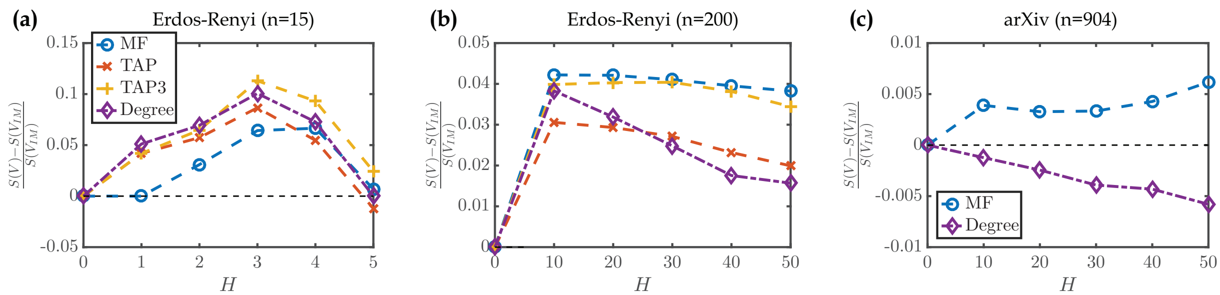

Figure 3: Comparison of the spread of influence under the linear threshold model for different greedy algorithms. For each network, we ensure $\sum_j J_{ij} \leq 1/2$ and we average over many draws of the initial bias $\{b_i^0\} \sim \mathcal{U}[-1/2, 1/2]$. (a) An Erdös-Rényi network with $n = 15$ nodes. (b) An Erdös-Rényi network with $n = 200$ nodes. (c) A collaboration network of $n = 904$ physicists on the arXiv. The benchmark in all panels is IM.

The negative bias $\theta_i = -b'_i$ is referred to as the threshold of i, representing the amount of influence i must receive from her neighbors to become active.

We include a stochastic influence ϵ for each node at every iteration t of the LT dynamics, representing natural fluctuations in the influence on each node over time. If ϵ is drawn from a logistic distribution, then these stochastic dynamics are equivalent to Glauber Monte Carlo dynamics (Newman and Barkema 1999)

$$P\big(\sigma_i^{(t+1)}\big|\boldsymbol{\sigma}^{(t)}\big) = \Big(1 + e^{-\frac{2}{T}(\sum_j J_{ij}\sigma_j^{(t)} + b_i)}\Big)^{-1}, \qquad (13)$$

where T parameterizes the variance of ϵ. Allowing the system time to equilibrate, the statistics of the Glauber dynamics follow the Boltzmann distribution (noting that we have taken $T = 1$ in Eq. (1)), simulating an Ising model. Furthermore, it is clear to see that we recover the deterministic LT dynamics in the limit of zero fluctuations $T \to 0$. Thus, the Ising model represents a natural generalization of the LT model to settings with stochastic fluctuations in the influence. We emphasize that, because maximum entropy models have demonstrated tremendous ability in quantitatively describing a wide range of real-world systems (Schneidman et al. 2006; Ganmor, Segev, and Schneidman 2011; Bialek et al. 2012; Phillips, Anderson, and Schapire 2006; Stein, Marks, and Sander 2015; Kapur 1989), understanding how these systems react to external influence is a significant endeavor in and of itself, and this goal should fundamentally be viewed as running adjacent to, as opposed to in conflict with, the existing viral influence maximization literature.

Finally, we note that most applications of the LT model to influence maximization impose the constraint $\sum_j J'_{ij} \leq 1$ ($\sum_j J_{ij} \leq 1/2$ in Ising notation) and assume that the bias b'_i is drawn uniformly from $[0, 1]$ ($b_i \sim \mathcal{U}[-1/2, 1/2]$) at the beginning of each simulation. Randomly selecting b_i at the beginning of each simulation is meant to represent uncertainty in the nodes' biases, which is fundamentally distinct from randomizing b_i at *each iteration* to represent natural fluctuations in the biases over time. Indeed, in the following experiments we include both sources of randomness, making our model equivalent to the so-called random-field Ising model, which shows similar behavior to a spin glass (Nattermann 1997).

Experimental Evaluation

We experimentally evaluate the performance of our greedy algorithm under various orders of the Plefka expansion. We also provide the first comparison between Ising influence algorithms and the traditional greedy influence maximization algorithm in (Kempe, Kleinberg, and Tardos 2003). As is usually assumed in traditional influence maximization, we scale the interactions such that $\sum_j J_{ij} \leq 1/2$. Furthermore, we draw the initial bias on each node from a uniform distribution $b_i \sim \mathcal{U}[-1/2, 1/2]$ and average over many such draws.

To fairly compare the Ising and linear threshold algorithms, we divide the experiments into two classes: the first is evaluated with respect to the total activity $\mathcal{M}$ under the Ising model, while the second class of experiments evaluates the spread S resulting from each choice of nodes under the linear threshold model. We denote the greedy algorithm in (Kempe, Kleinberg, and Tardos 2003) by IM. We also compare with the heuristic strategy of choosing the top H nodes with the highest degrees in the network.

Comparison Under the Ising Model. We first compare the different greedy algorithms under the Ising model. In Figs. 2(a-b), we use the Ising greedy algorithm GI with the third-order approximation TAP3 as the benchmark. In both Erdös-Rényi networks, we find that GI with TAP3 slightly outperforms TAP and MF, while all three Ising-based algorithms significantly out-perform the linear-threshold-based algorithm IM and the degree heuristic. Since TAP3, TAP, and MF all perform within 5% of one another, in Fig. 2(c) we use GI with the MF approximation as the benchmark. In the arXiv network, we find that GI with MF significantly outperforms both LT and the degree heuristic. These results demonstrate the practical importance of accurately modeling the stochastic noise in the system. We point out that the total activity $\mathcal{M}$ is calculated exactly in Fig. 2(a) using the Boltzmann distribution and estimated in Figs. 2(b-c) using Monte Carlo simulations.

Comparison Under the Linear Threshold Model. We now compare the algorithms under the linear threshold model. In all of Figs. 3(a-c), we use the LT greedy algorithm IM as the benchmark and exactly compute the spread of influence under the LT model. Surprisingly, in both of the Erdös-Rényi networks in Figs. 3(a-b), all of the Ising-based algorithms and the degree heuristic out-perform IM. In particular, the third-order approximation TAP3 achieves close to the largest spread in both networks. In the arXiv network in Fig. 3(c), the Ising-based algorithm continues to slightly out-perform IM, while IM out-performs the degree heuristic.

The success of the Ising-based algorithms is surprising, since they are all attempting to maximize a fundamentally different objective from influence spread under LT. We suspect that the strong performance might be the result of the Ising model performing a type of simulated annealing, similar to recent techniques proposed in (Jiang et al. 2011); however, an investigation of this hypothesis is beyond the scope of the current paper.

Conclusions

Maximum entropy models such as the Ising model have recently been used to quantitatively describe a plethora of biological, physical, and social systems. Given the success of the Ising model in capturing real-world systems, including populations of neurons in the brain and networks of interacting humans, understanding how to control and optimize the large-scale behavior of complex systems is of fundamental practical and scientific interest, with applications from guiding healthy brain development (Goddard, McIntyre, and Leech 1969) and intervening to alleviate diseased brain states (Goddard 1967) to anticipating trends in financial markets (Mantegna and Stanley 1999) and preventing viral epidemics (Pastor-Satorras and Vespignani 2001).

Here, we study the problem of maximizing the total activity of an Ising network given a budget of external influence. In the continuous setting where one can tune the influence on each node, we present a series of increasingly-accurate gradient ascent algorithms based on an approximation technique known as the Plefka expansion. In the discrete setting where one chooses a set of influential nodes, we propose a greedy algorithm and present novel conditions for when the objective is submodular.

Future Work. Given the novelty of this problem and the recent surge in the use of the Ising model, there are many promising directions to pursue. One direction is to consider a more general control problem in which the controller aims to shift the Ising network into a desired state instead of simply maximizing the activity of all nodes.

Another direction is to consider data-based optimization, wherein the optimizer is only aware of the past activity of the system (Goyal, Bonchi, and Lakshmanan 2011). Since the Ising model is mathematically equivalent to a Boltzmann machine (Coughlin and Baran 1995), one could adapt state-of-the-art methods from machine learning to approach this problem.

Finally, given the experimental success of the Ising-based greedy algorithms under the linear threshold model, an obvious extension of the current work should look into possible explanations. We suspect that a closer comparison with simulated annealing techniques in (Jiang et al. 2011) might provide the answer.

Acknowledgements

We thank Haim Sompolinsky and Eric Horsley for their helpful discussions. We acknowledge support from the U.S. National Science Foundation, the Air Force Office of Scientific Research, and the Department of Transportation.

References

Ackley, D. H.; Hinton, G. E.; and Sejnowski, T. J. 1985. A learning algorithm for boltzmann machines. *Cog. Sci.* 9(1):147–169.

Aurell, E., and Ekeberg, M. 2012. Inverse ising inference using all the data. *Phys. Rev. Lett.* 108(9):090201.

Bialek, W.; Cavagna, A.; Giardina, I.; Mora, T.; Silvestri, E.; Viale, M.; and Walczak, A. M. 2012. Statistical mechanics for natural flocks of birds. *PNAS* 109(13):4786–4791.

Bolthausen, E. 2014. An iterative construction of solutions of the tap equations for the sherrington–kirkpatrick model. *Comm. Math. Phys.* 325(1):333–366.

Brush, S. G. 1967. History of the lenz-ising model. *Rev. Mod. Phys.* 39(4):883.

Chen, W.; Wang, C.; and Wang, Y. 2010. Scalable influence maximization for prevalent viral marketing in large-scale social networks. In *SIGKDD*, 1029–1038. ACM.

Chen, W.; Wang, Y.; and Yang, S. 2009. Efficient influence maximization in social networks. In *SIGKDD*, 199–208. ACM.

Coughlin, J. P., and Baran, R. H. 1995. *Neural computation in hopfield networks and boltzmann machines*. University of Delaware Press.

Domingos, P., and Richardson, M. 2001. Mining the network value of customers. In *KDD*, 57–66. ACM.

Duchi, J.; Shalev-Shwartz, S.; Singer, Y.; and Chandra, T. 2008. Efficient projections onto the l 1-ball for learning in high dimensions. *ICML* 272–279.

Galam, S. 2008. Sociophysics: a review of galam models. *Int. J. Mod. Phys. C* 19(3):409–440.

Ganmor, E.; Segev, R.; and Schneidman, E. 2011. Sparse low-order interaction network underlies a highly correlated and learnable neural population code. *Proc. Natl. Acd. Sci.* 108(23):9679–9684.

Goddard, G. V.; McIntyre, D. C.; and Leech, C. K. 1969. A permanent change in brain function resulting from daily electrical stimulation. *Exp. Neurol.* 25(3):295–330.

Goddard, G. V. 1967. Development of epileptic seizures through brain stimulation at low intensity. *Nature* 214(5092):1020–1021.

Goyal, A.; Bonchi, F.; and Lakshmanan, L. V. 2011. A data-based approach to social influence maximization. *VLDB Endowment* 5(1):73–84.

Griffiths, R.; Hurst, C.; and Sherman, S. 1970. Concavity of magnetization of an ising ferromagnet in a positive external field. *J. Math. Phys.* 11:790.

Griffiths, R. B. 1967. Correlations in ising ferromagnets. i. *J. Math. Phys.* 8(3):478–483.

Hillis, W. D. 1988. Intelligence as an emergent behavior; or, the songs of eden. *Daedalus* 175–189.

Jaynes, E. T. 1957. Information theory and statistical mechanics. *Phys. Rev.* 106(4):620.

Jiang, Q.; Song, G.; Cong, G.; Wang, Y.; Si, W.; and Xie, K. 2011. Simulated annealing based influence maximization in social networks. In *AAAI*, volume 11, 127–132.

Kapur, J. N. 1989. *Maximum-entropy models in science and engineering*. John Wiley & Sons.

Kempe, D.; Kleinberg, J. M.; and Tardos, E. 2003. Maximizing the spread of influence through a social network. In *KDD*, 137–146. ACM.

Leskovec, J.; Krause, A.; Guestrin, C.; Faloutsos, C.; VanBriesen, J.; and Glance, N. 2007. Cost-effective outbreak detection in networks. In *KDD*, 420–429. ACM.

Lévy, P. 1997. *Collective intelligence*. Plenum/Harper Collins New York.

Lezon, T. R.; Banavar, J. R.; Cieplak, M.; Maritan, A.; and Fedoroff, N. V. 2006. Using the principle of entropy maximization to infer genetic interaction networks from gene expression patterns. *PNAS* 103(50):19033–19038.

Lynn, C. W., and Lee, D. D. 2016. Maximizing influence in an ising network: A mean-field optimal solution. *NIPS*.

Lynn, C. W., and Lee, D. D. 2017. Statistical mechanics of influence maximization with thermal noise. *EPL*.

Lynn, C. W.; Papadopoulos, L.; Lee, D. D.; and Bassett, D. S. 2017. Surges of collective human activity emerge from simple pairwise interactions. *Preprint*.

Mantegna, R. N., and Stanley, H. E. 1999. *Introduction to econophysics: correlations and complexity in finance*. Cambridge University Press.

Mataric, M. J. 1993. Designing emergent behaviors: From local interactions to collective intelligence. In *SAB*, 432–441.

Mora, T.; Walczak, A. M.; Bialek, W.; and Callan, C. G. 2010. Maximum entropy models for antibody diversity. *PNAS* 107(12):5405–5410.

Namatame, A.; Kurihara, S.; and Nakashima, H. 2007. *Emergent Intelligence of Networked Agents*, volume 56. Springer.

Nattermann, T. 1997. Theory of the random field ising model. *Spin glasses and random fields* 12:277.

Nesterov, Y. 2013. *Introductory lectures on convex optimization: A basic course*, volume 87.

Newman, M., and Barkema, G. 1999. *Monte Carlo Methods in Statistical Physics*. New York: Oxford University Press.

Pastor-Satorras, R., and Vespignani, A. 2001. Epidemic spreading in scale-free networks. *Phys. Rev. Lett.* 86(14):3200.

Phillips, S. J.; Anderson, R. P.; and Schapire, R. E. 2006. Maximum entropy modeling of species geographic distributions. *Ecol. Modell.* 190(3):231–259.

Plefka, T. 1982. Convergence condition of the tap equation for the infinite-ranged ising spin glass model. *J. Phys. A* 15(6).

Schneidman, E.; Berry, M. J.; II, R. S.; and Bialek, W. 2006. Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature* 440(7087):1007.

Stein, R. R.; Marks, D. S.; and Sander, C. 2015. Inferring pairwise interactions from biological data using maximum-entropy probability models. *PLoS Comp. Bio.* 11(7).

Sylvester, G. S. 1983. Magnetization concavity in ferromagnets. *J. Stat. Phys.* 33(1):91–98.

Yedidia, J. 2001. *Advanced mean field methods: Theory and practice*.