

Information Diffusion of Topic Propagation in Social Media

Shahin Mahdizadehghadam¹, Han Wang¹, Hamid Krim¹, and Liyi Dai²

¹Department of Electrical and Computer Engineering,
North Carolina State University, Raleigh, NC

²Army Research Office, RTP, Raleigh NC

Abstract—Real-world social and/or operational networks consist of agents with associated states, whose connectivity forms complex topologies. This complexity is further compounded by interconnected information layers, consisting, for instance, documents/resources of the agents which mutually share topical similarities. Our goal in this work is to predict the specific states of the agents, as their observed resources evolve in time and get updated. The information diffusion among the agents and the publications themselves effectively result in a dynamic process which we capture by an interconnected system of networks (i.e. layered). More specifically, we use a notion of a supra-Laplacian matrix to address such a generalized diffusion of an interconnected network starting with the classical "graph Laplacian". The auxiliary and external input update is modeled by a multidimensional Brownian process, yielding two contributions to the variations in the states of the agents: one that is due to the intrinsic interactions in the network system, and the other due to the external inputs or innovations. A variation on this theme, a priori knowledge of a fraction of the agents' states is shown to lead to a Kalman predictor problem. This helps us refine the predicted states exploiting the error in estimating the states of agents.

Three real-world datasets are used to evaluate and validate the information diffusion process in this novel layered network approach. Our results demonstrate a lower prediction error when using the interconnected network rather than the single connectivity layer between the agents. The prediction error is further improved by using the estimated diffusion connection and by applying the Kalman approach with partial observations.

Index Terms—Multi-layer network, Information diffusion, Topic propagation, Computer networks.

I. INTRODUCTION

The emergence and rapid growth of the Internet have led to far greater access to and exchange of information between machines and among people. Social media such as Facebook, Twitter, and LinkedIn, have indeed connected billions of people across the planet, and have in addition almost trivialized information and resource exchange. As a physical process, the information flow—through its rather extensive connectivity and resulting frequent exchanges among the nodes, as connecting entities—is a pervasive diffusion. The

high activity over a typical network entails active dynamics, and hence large variations in content. The states of agents (nodes) in such a network are consequently actively altered by the diffusion of information, according to associated dynamical parameters of the network. Throughout this work, we use the general term "agent" to refer to users in social networks, bloggers, authors, or any individual who produces information content, and is networked with others. Agents are usually/invariably interested in different topics irrespective of the structure of the network. Bloggers, for example, may write about specific topics, while different authors may only publish articles about their own research areas.

The interest of a given agent may, also, evolve over time, as a result of diffusion of information from other agents or sources. This work is, to the best of our knowledge, first in addressing this problem in its full generality thus, in particular, we propose *layered and interconnected networks* to beyond a two-layer network, as has appeared in the literature [9]. This in turn allows diffusion to take place over many layers (a truly *multi-layer diffusion network*).

To proceed, we note that the diffusion phenomenon in a network and our later formulation will also be useful in predicting the future state of agents in the network. Specifically, information diffusion process on the topical states of agents is further clarified with the following fundamental working hypotheses:

1) *Information and State of Agents:*

To successfully model the information diffusion, it is first imperative to carefully define the type of dynamics which characterize the diffusion process. Many existing diffusion-based methods have associated characteristics such as infections, epidemics, diseases, viruses or contagions which can spread over the network. For completeness, we briefly describe some of these works: in Susceptible-Infected-Susceptible (SIS) models [24,25,31] the infection can spread from any infected agent to its neighbors under certain conditions and sets of probabilities, and the agents can either be in susceptible or infected states. In this case, an agent in an infected state, will switch to a susceptible state over time, with some probability p . However, in Susceptible-Infected-Recovered (SIR) models [6,16,21,34] the agents will switch to the recovered state, and may not catch the infection again.

smahdiz@ncsu.edu
hwang42@ncsu.edu
ahk@ncsu.edu
liyi.dai.civ@mail.mil

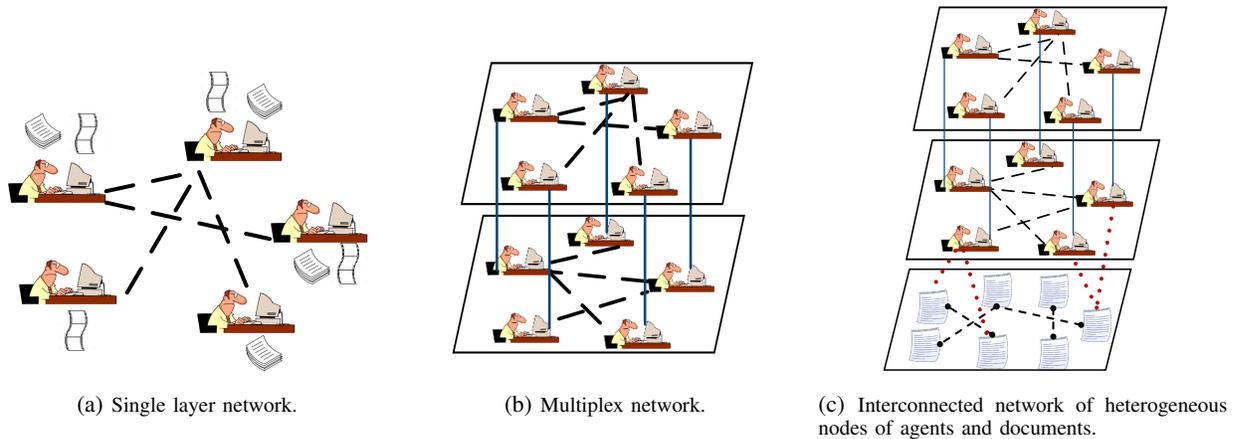


Fig. 1: a) A single layer network between bloggers. Dashed edges show connections between bloggers. b) A multiplex network between bloggers. The top layer is based on the hyperlink inter-connectivity of the bloggers, and the bottom layer is the friendship network between the bloggers. The straight inter-layer edges are showing that the bloggers are the same people in both layers. c) An interconnected network of heterogeneous nodes of agents and documents. The dotted edges are showing which blogger (agent) has produced which document. A similar set of dotted inter-layer edges are present from the top layer to the bottom layer which has not been depicted in the picture.

In addition, one commonly defines the state of an agent as active or a passive [2, 33]. The active agents will influence their neighbors to activate/trigger them.

Despite the subtle differences in these models, the state space of the agents is usually binary indicating influence from a certain information. While these types of diffusions may capture some dynamics, they present serious limitations in any practical network settings. A case in point is that of a diffusion process due to a blogger’s read of other blogs in a network, and whose state/opinion changes as a result. This in turn, changes the dynamics of the network. In Instagrams, users may “repost” each other’s images (with or without modification of the original image), thus implying that the information diffusing through the network may come from text, image, video, and other types of documents.

2) The connectivity model:

Information diffusion over a network is intimately related to the connectivity structure. While the connectivity model describes nodes with a potential to influence/(or be influenced by) others, it may assign a quantitative value on the relation of the agents defined by their mutual influence. Graphs are most commonly used to model the connectivity among agents who are represented by nodes, and whose connection is depicted by edges.

In the aforementioned “bloggers” example, connectivity between bloggers highlights the fact that they read each other’s blog. Fig. (1a) shows a single layer connectivity structure between the bloggers; but various structures between them are also possible. We may, for instance, consider the information about the hyperlink connectivity as well as the friendship network between the bloggers. Fig. (1b) depicts a two-layer connectivity structure (multiplex with two layers) between bloggers.

In this work we define the state of each agent as a feature vector. Specifically, we will consider the state of each agent as a topic vector, to reflect the extent to which an agent is

associated with specific topics. Our goal in this endeavor is to also provide the evolution of the topic vector of each agent over time. The goals and contributions in this paper are: creating a multi-layer diffusion network among agents and information contents, estimating the future state of agents, learning the structure of the supra-Laplacian matrix from previous diffusion history, and predicting the future states assuming that a partial observation of the states of agents is available.

1) We go beyond the two-layer case to account for topical connectivity among the content nodes (sources of data characteristic of the information permeating the interconnected network) thus yielding a multi-layer network. In our example, blogs consist of a set of documents generated by bloggers over time, and as illustrated in Fig. (1c). These document sets may share some topical similarities (in spite of their distinct blogger sources). Structuring these relational properties into a network of document sets, introduces additional connecting edges between agents and document sets, as well as among agents. These additional paths will diffuse information at a secondary degree, specifically, an absence of a direct connection between two bloggers, does not exclude one blogger getting updated on another blogger, and this on account of topical similarities between documents, they are associated with. We will refer to this complex structure of connected networks, as an *interconnected network* of heterogeneous nodes, with advantages further discussed in Section III. This structure will help us understand the process in which the author’s topical interests are changing from a topic to another, or why different topics are getting different amounts of attention by users in online social networks.

2) While the current states of agents represent their engagement in different topics, they will evolve over time to reflect new documents being produced, hence updating the agents’ states. We propose to estimate the state of the agents

at time t_1 , knowing their state at time $t_0 < t_1$ with the help of the interconnected network. This is accomplished by considering a multi-layer diffusion network among the agents and documents.

3) In many cases, the actual diffusion among the agents is different from the fixed connectivity which we observe in the network (e.g, for a given friendship network between bloggers, a connection between two friend bloggers does not necessarily imply an influence of one on the other.). By observing diffusion related patterns over a short period of time, we particularly attempt to understand such a *static* influence network (reasonably for a short time interval), which has the presumably close connection to agents involvement in some common interest, with the capacity of leading to a common consensus. To that end, we learn the structure of the supra-Laplacian matrix of the underlying network of influence, which we subsequently use that in our prediction phase.

4) In certain online social networks, one often encounters a set of conservative users with strict privacy policy and a fraction of hub nodes with public information, and this clearly alters the previously designed state prediction problem. We exploit the flexibility of the Kalman filter to address the problem, assuming that a partial observation of the states of agents is available. The Kalman filter enables us to refine the predicted states by exploiting the prediction error in the available states.

This paper is organized as follows: In Section II we discuss some background and related work to our contribution. Section III describes our basic formulation of the problem. We propose our new approach in Section IV, and present substantiating experimental results in Section V. We provide some concluding remarks in Section VI.

II. RELATED WORKS

In recent years, a number of studies of real-networks have focused on epidemic spreading in multiplex networks. In [32], two epidemics spreading in a two-layered multiplex network was addressed. While agents in the two-layered multiplex network are the same, their connectivity structures are different. In the state transition diagram of the model, each agent can be infected either by the first epidemic (I_1), or by the second epidemic (I_2), or possibly susceptible state (S). The suggested model is in the form of SI_1I_2S . The infected agent by the first epidemic should first get into a susceptible state to later get infected by the second epidemic. Also, thresholds determining the epidemics persistence, are defined. In [4], the authors study a similar environment (SI_1SI_2S) in a two-layered multiplex network. The work focused on the long-term extinction, coexistence and absolute dominance of the two epidemics with a different definition of dominance than that of [32]. In [9], the authors discuss the diffusion dynamics in a two-layered multiplex network. The authors introduce supra-Laplacian matrix, and used a perturbation analysis to study the effects of changes in the spectral properties of single layer on the spectral

properties of the whole network. Similarly in [8], the authors study the spreading of two different processes in a multiplex network. Their work mostly focused on the impact of the degree correlation of the two layers on the epidemic properties of the processes. The state space they considered, similar to many others was of binary nature (infected or not-infected) to model the spreading processes.

We can classify the information diffusion process models into three major groups, probabilistic models, thermodynamic models, and counting models. NETINF [10], NETRATE [22] and INFOPATH [11] are the probabilistic models which infer the underlying diffusion network between information sources using consecutive hit times of the nodes by a specific cascade. The diffusion network was the result of solving an optimization problem which is computationally very expensive (super-exponential). The main idea behind the thermodynamic models [7, 9, 19, 27] is that heat will propagate from a warmer region to a colder region or gas will move from the region with higher density to the region with lower density. Modeling the information as heat or gas, we can write the rate at which information is changing in agent i as: $\frac{d\psi_i}{dt} = D \sum_j \mathbf{A}(i, j)(\psi_j - \psi_i)$. Where $\psi_i(t)$ is the state of the i^{th} agent at time t , D is the diffusion constant which is the amount of information passing from an agent to another agent in a small interval of time, and $\mathbf{A}(i, j)$, the so-called adjacency matrix, reflects the connectivity between agents i and j . The counting models [12, 18, 26] form counting processes to find the number of nodes in each group of susceptible or infected nodes. Assuming that each agent has β contacts with other agents per unit time, and in each contact of an infected agent with a susceptible node, the diseases will definitely spread, the overall rate of new infections is $\beta SX/n$, where n is the total number of nodes. X and S are the number of infected and susceptible individuals respectively. We therefore, can write the rate of change of X and S as $\frac{dX}{dt} = \beta \frac{SX}{n}$ and $\frac{dS}{dt} = -\beta \frac{SX}{n}$.

In this paper we address following limitations from the aforementioned information diffusion models.

A) Almost all existing information diffusion models are based on the contagion principle, using an abstract term or a clear data tag. Specifically, many information diffusion models in social networks, for example, consider a specific "Hashtag" or repost as contagion [7, 9, 10, 27]. We would, thus, need to advance existing information diffusion models, if one were interested in, tracking the way bloggers influence each other, or how articles and related research fields vary over time, or parsing through very large multi-modal data sets.

B) A few information diffusion models [9, 27] have recently considered multiplex connectivity models. The overwhelming majority of these works model the connectivity by just using simple graphs. Agents may, however, display several forms of connectivity, through multiple online social networks, or other social clubs, thus making simple graph models highly biased for capturing the real world complexity of information diffusion models.

C) Existing models are layered by lack of considering external perturbations/inputs. Specifically, considering the bloggers' example, the documentation evolves with time, and so do the nodes of the network, hence impacting the mutual influence the agents have on each other. This is indeed key to one's ability to proceed to any future state prediction on the network. In the present work, we model each agent state vector as a topic vector, effectively as a topic proportion associated with the agent's own documents. In so doing, our solution predicts the users' future topic vectors by considering their interactions inside the interconnect network agents, as well as those of the documents. The latter network is allowed to vary over time on account of both the diffusion effect as well as that of the external input as new documents are added.

Our work is similar to the research in [9], we however, consider the interconnected network of heterogeneous nodes of documents and agents (Fig. 1c) rather than the simpler multiplex network (Fig. 1b) in [9]. We will also generalize the scalar state space of agents to a higher dimensional state space of topic proportions. Furthermore, we will consider an open system, and we study the external effect on each agent from the agents or documents which have not been captured by the diffusion in the network.

III. PROBLEM FORMULATION

We represent agents within a social network setting, by nodes on a graph, and their connectivity by edges between them. The connectivity may reflect friendship, co-authorship, or other forms of behavioral similarities. If we consider a set V_A of N agents, we can imagine M_A different graphs $G_A^{(m)} = (V_A^{(m)}, E_A^{(m)})$, $m = 1, \dots, M_A$ based on the M_A different connectivity structures among these N agents. We will use throughout, the term "agent-layer" to refer to the network comprised of communicating agents. In Fig. (2) we can see two agent-layers. We note in this case, that the agents are the same in both layers, and we hence have a direct one to one connections between them (the blue edges in Fig. (2)).

Agents are usually assumed to possess some data or to produce information in their embedding network. If, for instance, the agents were bloggers or social network users, they would be producing blog-documents or their online accounts, if they were researchers, they would be turning out scientific papers. These sets of documents, may also mutually store topical similarities which will play a role in the overall information diffusion throughout the network. Of central interest in our work, is the evolution of the topical states of the agents, as the information diffusion takes place. Latent Dirichlet Analysis (LDA) [1], Latent Semantic Indexing (LSI) [5] and several other topic modeling algorithms afford us the ability to represent any text document as state vectors of length T . T is an arbitrarily chosen number of topics (dimension) in a collection of text documents. Considering the documents as vectors in a T

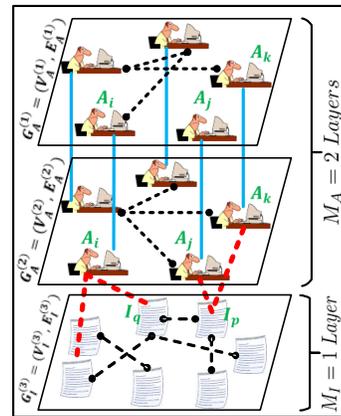


Fig. 2: The interconnected network \mathcal{I} (Eqn. (2)) with two agent-layers and one information-layer. A set of inter-layer edges (similar to the red lines) are present from the top layer to the bottom layer which have not been depicted in the picture for simplicity.

dimensional vector space, we can build a similarity network between these vectors using their euclidean inter-distances, as well as a neighborhood criterion between them (e.g. K -Nearest Neighbors algorithm, Epsilon-Neighborhood). We consider a set V_I of S documents. The subscript I specifies the information nature of the network and of its members (i.e, it pertains to documents in this case), with M_I different graphs $G_I^{(m)} = (V_I^{(m)}, E_I^{(m)})$, $m = M_A + 1, \dots, M_A + M_I$, each describing a different connectivity structure among S documents (the bottom layer in Fig. (2)). Note that similarly efficient tools (e.g. ISOMAP [28], Principal Component Analysis (PCA) [13], Multi-Dimensional Scaling (MDS) [29]) for dimension reduction are available for multi-modal documents (images, videos, ...), to possibly represent them as separate networks. In this work, we refer to the network among the documents as the information-layer. Table I lists all notations used in this paper along with their definitions.

By adopting an adapted notation of [15], we represent the document p by I_p , and the agent i by A_i . The red inter-layer edges ($E_{A,I}$) in Fig. (2) relates agents to their respective documents (publisher-publication network). The information of document I_p is denoted by a T dimensional vector $\mathbf{x}_{I_p} \in \mathbb{R}^T$. The state of agent A_i in layer m , is defined as the average of the T dimensional vectors of their associated documents:

$$\mathbf{x}_{A_i}^{(m)} = \frac{1}{|\mathcal{N}_I(A_i)|} \sum_{p \in \mathcal{N}_I(A_i)} \mathbf{x}_{I_p}, \quad (1)$$

where $\mathcal{N}_I(A_i)$ is the set of documents produced by agent A_i or more formally, the neighbor set of agent A_i in the information-layer, which $|\mathcal{N}_I(A_i)|$ is the cardinality of that set. The interconnected graph in Fig. (2) may thus be denoted by:

$$\mathcal{I} = \left((V_A^{(1)} \cup V_A^{(2)} \cup V_I^{(3)}), (E_A^{(1)} \cup E_A^{(2)} \cup E_I^{(3)} \cup E_{A,A}^{(1,2)} \cup E_{A,I}^{(1,3)} \cup E_{A,I}^{(2,3)}) \right). \quad (2)$$

TABLE I: Terminology

Symbol	Definition	Symbol	Definition
A, I	Agents & documents resp.	$\mathbf{W}_A^{(k)}$	Intra-layer connectivity matrix of agents in layer k .
A_i, I_p	Agent i & documents p resp.	$\mathbf{W}_{A,I}^{(k,m)}$	Inter-layer conn. of layer k to m among agents and docs.
$G_A^{(m)}, G_I^{(m)}$	Intra-layer graphs of layer m .	$D_A^{(k)}$	Intra-layer diffusion constant of the agents in layer k .
\mathcal{I}	The interconnected graph.	$D_A^{(k,l)}$	Diffusion constant of agents from layer k to l .
T	Length of the topic vectors.	$D_{A,I}^{(k,m)}$	Diffusion const. of layer k to m . among agents and docs.
$\mathbf{L}^{(m)}$	Laplacian matrix of layer m .	N, S, P	# of agents, documents & all nodes resp.
$\mathbf{x}_{A_i}^{(m)}$	State of agent i in layer m .	\mathcal{L}	Supra-Laplacian matrix.
\mathbf{x}_{I_p}	State of document p .	\mathbf{X}	A $P \times T$ matrix consisted of the state of all nodes.

With this inter-connected formulation, and in analyzing an associated information diffusion process, we address in this paper,

- The ability to achieve a diffusion of information between agents via multiple connectivity structures. The same set of agents may have different intra-layer connectivity at different layers;
- The ability to model networked documents in a separate network layer, will enable us to consider the similarities, and their evolution overtime between the documents as an information diffusion medium. For instance, agents A_i and A_j in Fig. (2) are not connected via any paths through the top two agent-layers $G_A^{(1)}$, and $G_A^{(2)}$. There is, however, a path between these two agents by way of the similarity of their documents I_p and I_q in $G_I^{(3)}$. The intuition supporting the interaction between A_i and the A_j in the blogger's example, follows from: blogger A_i noticing I_p document's similarity (written by blogger A_j) to his/her I_q document, would read the document and being influenced, as a result.
- The preservation of the conventional information diffusion structures such as the co-authorship network in the introduced interconnected network model (e.g the agents A_j and A_k who have collaborated to produce the I_p).

IV. THE PROPOSED METHOD

In this section we address the information diffusion across agents as a result of topic adoption and adaptation, as well as external topic additions. To that end, we next consider a closed interconnected network with no additions. In Section IV-B, we consider an open interconnected network and account for an innovation injection. In Section IV-C, we define the estimation of the supra-Laplacian matrix using learning data. In Section IV-D, we further refine the predicted state of the nodes using Kalman filtering. In Section IV-E, we analyze the effect of a weak inter-layer connection on the over-all diffusibility of the interconnected network.

A. Closed system: Information diffusion in a heterogeneous network

In closed systems, all changes in the states of agents are a result of interaction of the agents in the network. In a single

layer diffusion process [7, 9, 19, 27], an agent state maybe formalized as:

$$\frac{d\mathbf{x}_{A_i}}{dt} = D \sum_{j=1}^N \mathbf{W}(i, j)(\mathbf{x}_{A_j} - \mathbf{x}_{A_i}), \quad (3)$$

where $\frac{d\mathbf{x}_{A_i}}{dt}$ is the i^{th} agents' topic vector change over time. $\mathbf{W}(i, j)$ reflects the connectivity status between agents i and j , while D is the diffusion constant, or the fractional amount of information passing from j to i in a small time interval. By further simplification of Eqn. (3), we may write $\frac{d\mathbf{X}_A}{dt}$ as follows:

$$\frac{d\mathbf{X}_A}{dt} + D\mathbf{L}\mathbf{X}_A = 0, \quad (4)$$

where \mathbf{X}_A is an $N \times T$ matrix, where row i is denoted by the row vector $\mathbf{x}_{A_i}^T$; \mathbf{L} is an $N \times N$ graph Laplacian matrix (i.e., $\mathbf{L} = \mathbf{K} - \mathbf{W}$, \mathbf{K} being the diagonal matrix of the nodes' degree).

Independent of the nodes in the network being agents or documents, we can state the following,

Proposition 1. *We can generally write the supra-Laplacian matrix of an M layer multiplex network with N nodes in each layer as $\mathcal{L} = \mathcal{L}_L + \mathcal{L}_I$. Where \mathcal{L}_L is the supra-Laplacian matrix of the intra-layer connectivity and \mathcal{L}_I is the supra-Laplacian matrix of the inter-layer connectivity. \mathcal{L}_L may be in turn, written as direct sum of the Laplacian matrices of the independent intra-layer connectivities:*

$$\mathcal{L} = \mathcal{L}_L + \mathcal{L}_I, \quad (5)$$

$$\mathcal{L}_L = \bigoplus_{\alpha=1}^M D^{(\alpha)}\mathbf{L}^{(\alpha)} = \begin{bmatrix} D^{(1)}\mathbf{L}^{(1)} & & & \\ & D^{(2)}\mathbf{L}^{(2)} & & \\ & & \ddots & \\ & & & D^{(M)}\mathbf{L}^{(M)} \end{bmatrix}. \quad (6)$$

The inter-layer supra-Laplacian can be written as $\mathcal{L}_I = \sum_{\alpha=1}^M (\mathcal{K}_I^\alpha - \mathcal{W}_I^\alpha)$, where the \mathcal{K}_I^α is the diagonal inter-layer degree matrix of layer α , showing the inter-layer degree of the nodes in layer α and the \mathcal{W}_I^α is the inter-layer connectivity matrix of the nodes in layer α with the nodes in the other layers. The \mathcal{K}_I^α and \mathcal{W}_I^α are formally defined in Eqns. (7 and 8) respectively:

$$\mathcal{K}_I^\alpha = \mathbf{e}_{(\alpha, \alpha)} \otimes \left(\sum_{\beta=1, \beta \neq \alpha}^M D^{(\alpha, \beta)} \mathbf{K}^{(\alpha, \beta)} \right), \quad (7)$$

$$\mathcal{L} = \begin{bmatrix} D^{(1)}\mathbf{L}^{(1)} + D^{(1,2)}\mathbf{I} + D^{(1,3)}\mathbf{K}^{(1,3)} & -D^{(1,2)}\mathbf{I} & -D^{(1,3)}\mathbf{W}^{(1,3)} \\ -D^{(2,1)}\mathbf{I} & D^{(2)}\mathbf{L}^{(2)} + D^{(2,1)}\mathbf{I} + D^{(2,3)}\mathbf{K}^{(2,3)} & -D^{(2,3)}\mathbf{W}^{(2,3)} \\ -D^{(3,1)}\mathbf{W}^{(3,1)} & -D^{(3,2)}\mathbf{W}^{(3,2)} & D^{(3)}\mathbf{L}^{(3)} + D^{(3,1)}\mathbf{K}^{(3,1)} + D^{(3,2)}\mathbf{K}^{(3,2)} \end{bmatrix}. \quad (11)$$

$$\mathbf{W}_I^\alpha = \sum_{\beta=1, \beta \neq \alpha}^M (\mathbf{e}_{(\alpha, \beta)} \otimes (D^{(\alpha, \beta)} \mathbf{W}^{(\alpha, \beta)})), \quad (8)$$

where $\mathbf{K}^{(\alpha, \beta)}$, is the diagonal matrix reflecting the degree of each node in the inter-layer connectivity between layer α and layer β , $\mathbf{W}^{(\alpha, \beta)}$ quantifies the inter-layer connectivity of the layer α nodes to the layer β nodes and $\mathbf{e}_{(\alpha, \beta)}$ is an all 0, $M \times M$, matrix with an only 1 element in (α, β) .

Proof. Considering the interconnected network case depicted in Fig. (2), with M_A and M_I layers of connectivity among the agents and the documents respectively, we can write,

$$\begin{aligned} \frac{d\mathbf{x}_{A_i}^{(k)}}{dt} &= D_A^{(k)} \sum_{j=1}^N \mathbf{W}_A^{(k)}(i, j) (\mathbf{x}_{A_j}^{(k)} - \mathbf{x}_{A_i}^{(k)}) \\ &+ \sum_{l=1}^{M_A} D_A^{(k, l)} (\mathbf{x}_{A_i}^{(l)} - \mathbf{x}_{A_i}^{(k)}) \\ &+ \sum_{m=M_A+1}^{M_A+M_I} D_{A, I}^{(k, m)} \sum_{p=1}^S \mathbf{W}_{A, I}^{(k, m)}(i, p) (\mathbf{x}_{I_p}^{(m)} - \mathbf{x}_{A_i}^{(k)}). \end{aligned} \quad (9)$$

The first term of Eqn. (9) represents the intra-layer diffusion of the information inside layer k , the second term accounts for the diffusion of the information between different agent-layers for agent i , while the third term describes the diffusion of information from different information-layers to agent i at layer k . $\mathbf{W}_A^{(k)}$ depicts the connectivity of agents within layer k , while $\mathbf{W}_{A, I}^{(k, m)}$ does that of agents in layer k and documents in layer m . The $D_A^{(k)}$ is the intra-layer diffusion constant of agents in layer k , while $D_A^{(k, l)}$ is the inter-layer diffusion constant of agents from layer k to agents in layer l , and $D_{A, I}^{(k, m)}$ is the inter-layer diffusion constant between agents in layer k and documents in layer m .

Further simplification of Eqn. (9) yields the following differential equation:

$$\frac{d\mathbf{X}}{dt} = -\mathcal{L}\mathbf{X}, \quad (10)$$

where \mathbf{X} is an $P \times T$ matrix. In case of an M layer multilayer network, rewriting the \mathcal{L} matrix will lead us to Eqn. (5). \square

Following is an example which further clears steps of the proof.

Example 1. In a three-layer interconnected network similar to Fig. (1c), \mathbf{X} is a $P \times T$ matrix ($P = 2N + S$) which represents the states of agents in the top two layers ($\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$), and of the topic vectors of the documents in the third layer ($\mathbf{X}^{(3)}$). Following the terminology in [9], we refer to \mathcal{L} as a supra-Laplacian matrix, by its ability to capture the diffusion in inter-connected network system. Assuming undirected graphs in each layer, and symmetric

diffusion constants ($D^{(k, m)} = D^{(m, k)}$), we can write for easier explanation, the supra-Laplacian matrix in case of a three layer-network (Fig. (2)), with two agent-layers and one information-layer as Eqn. (11). For this specific a 3 layer inter-connected network, we simplify our notation by dropping the subscripts (as in Eqn. (11), specifying the nature of a network layer) in favor of superscripts.¹ This hence makes matrix \mathbf{L}^m represent the Laplacian of the intra-layer connectivity matrix of layer m , \mathbf{I} the identity matrix, while \mathbf{K}^m is a diagonal matrix of node degree of layer m , and $\mathbf{K}^{m, n}$ is a diagonal matrix reflecting the degree of each node in the inter-layer connectivity between layer m and layer n . This hence yields Eqn. (11), which the first elements in the diagonal entries, form \mathcal{L}_L (Eqn. (6)), the other two elements in the diagonal entries form \mathcal{K}_I^α s (Eqn. (7)), and the rest (off-diagonal entries), form \mathcal{W}_I^α s (Eqn. (8)). \square

B. Open System Diffusion: Impact of External Effects

Much of the existing work in information diffusion models have a limited scope (of agents, documents, parameters) when predicting the future state of the nodes. More specifically, agent states may be varied by external sources which are not captured in the network, or by some agent actions which may even to some extent, conflict with the model prediction. to address this additional auxiliary input, we propose open system model as follows:

First, consider a single layer network (agent-layer) for simplicity, we can then model the rate of change in the state of node i as follows,

$$d\mathbf{x}_{A_i} = D \sum_{j=1}^N \mathbf{W}(i, j) (\mathbf{x}_{A_j} - \mathbf{x}_{A_i}) dt + \sigma_i d\mathbf{B}(t), \quad (12)$$

where $\mathbf{W}(i, j)$ quantifies the connectivity between agent i and j , and D is the diffusion constant reflecting the infinitesimal amount of information passing from j to i in a small interval of time, and $\mathbf{B}(t)$ is a $T \times T$ matrix, whose columns are T -dimensional vectors with components as independent standard Brownian motions of variances σ_i . Inspired by the Ornstein-Uhlenbeck (O.U.) process [30], Eqn. (12) describes the velocity of the topical-state of the nodes as a Brownian motion in presence of friction. In other words, to describe the uncertainty due to external effects, we proceed to view the whole system as a massive Brownian particle. The drift term (first term in right-hand side of Eqn. (12)), however, moves the velocity from a martingale state of $\sigma_i d\mathbf{B}(t)$ towards a consensus (captured by the drift term). In matrix form, Eqn. (12) may be written as,

$$d\mathbf{X}_A(t) = -D\mathbf{L}\mathbf{X}_A(t)dt + \Sigma_A d\mathbf{B}(t), \quad (13)$$

¹Layers 1-2 are the agent-layers, while layer 3 is the information-layer.

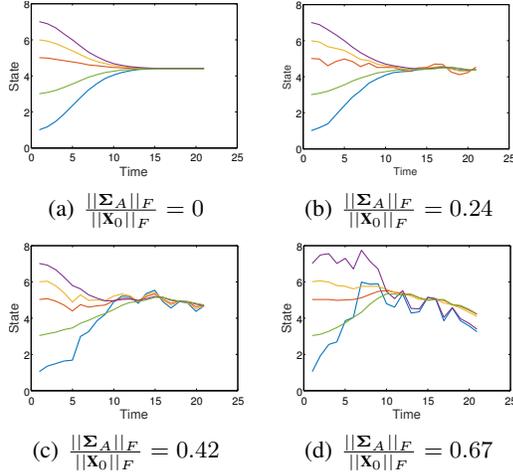


Fig. 3: The effect of different magnitude of Σ_A matrix on the state of agents.

where Σ_A is an $N \times T$ matrix, and each row shows the σ_i vector in Eqn. (12) for agent i . Fig. (3) shows numeric of examples of Eqn. (13) with $N = 5$ and $T = 1$. The Σ_A matrix adjusts the effect of the consensus and the Brownian motion in Eqn. (13). In this figure, we have used the term $\frac{\|\Sigma_A\|_F}{\|\mathbf{X}_0\|_F}$ (with $\mathbf{X}_0 = \mathbf{X}_A(0)$) to compare the effect of Σ_A magnitude matrix on the states of the agents. As the value of $\frac{\|\Sigma_A\|_F}{\|\mathbf{X}_0\|_F}$ increases the second term on the right-hand side of Eqn. (13) will have more influence on determining the state of the nodes and consequently, the state of the agents will be changed by higher uncertainty.

To extend to an interconnected setting, we can state the following statement: Given the states of nodes at time t_0 we can predict the states at time t_1 , $t_1 > t_0$ as follows:

$$\hat{\mathbf{X}}(t_1) = e^{-\mathcal{L}(t_1-t_0)}\mathbf{X}(t_0) + \int_{t_0}^{t_1} e^{\mathcal{L}(s-t_1+t_0)}\Sigma d\mathbf{B}(s), \quad (14)$$

where $e^{-\mathcal{L}(s-t_1+t_0)}$ is a matrix exponential and itself is a $P \times P$ matrix and Σ is $P \times T$ matrix. To this end, we rewrite Eqn. (13) for interconnected networks as,

$$d\mathbf{X}(t) = -\mathcal{L}\mathbf{X}(t)dt + \Sigma d\mathbf{B}(t). \quad (15)$$

The first term on the right-hand side of Eqn. (15) describes the network-level diffusion taking place among the nodes (mean reverting term), while the second depicts the global diffusion process, affecting nodes regardless of their interactions.

There are multiple ways to calculate the term $\int_{t_0}^{t_1} e^{\mathcal{L}\tau} d\tau$, one may use the basic definition of a matrix exponential, and calculate the numerical value of the integral, or instead use the Jordan form as $\mathcal{L} = \mathcal{M}\mathbf{J}\mathcal{M}^{-1}$ and $e^{\mathcal{L}\tau} = \mathcal{M}e^{\mathbf{J}\tau}\mathcal{M}^{-1}$.

Our proposed learning procedure will evaluate the diffusion constants in the supra-Laplacian matrix \mathcal{L} as well as the Σ matrix. To that end, we proceed to minimize the

Frobenius norm of the difference between \mathbf{X} and it's predict $\hat{\mathbf{X}}$, resulting from,

$$\arg \min_{\Sigma, D_1, \dots} g = \|\mathbf{X}(t_1) - \hat{\mathbf{X}}(t_1)\|_F. \quad (16)$$

Solving this optimization problem helps us decompose the predicted matrix into two main components on the right-hand side of Eqn. (14), the first term representing the interactions in the network, while the second quantifying the uncertainty which results from auxiliary inputs into the system.

C. Diffusion Network Estimation (Learning the Supra-Laplacian Matrix)

The supra-Laplacian matrix \mathcal{L} which we use in Eqn. (15) for state prediction, is a result of the network connectivity (refer to Eqn. (11)). In practice, hidden connections are pervasive, introducing uncertainty in the prediction, which are causing the information diffusion to require more than the predefined, explicit connections from the network. To that end, consider observations of $\mathbf{X}(t)$ over $t \in [0, t_1]$, denote $\bar{\mathbf{x}}(t) := \text{vec}(\mathbf{X}(t))$, the vectorization of $\mathbf{X}(t)$ to obtain a vector differential system in order to learn the supra-Laplacian matrix \mathcal{L} of Eqn. (15):

$$\dot{\bar{\mathbf{x}}}(t) = \Lambda \bar{\mathbf{x}}(t) + \bar{\mathbf{w}}(t), \quad 0 \leq t \leq t_1,$$

and we have $\Lambda = \mathbf{I}_T \otimes (-\mathcal{L})$, the Kronecker product of T -by- T identity matrix with $(-\mathcal{L})$ and $\bar{\mathbf{w}}(t)$ is the vectorization of $\mathbf{w}(t) = \Sigma \frac{d\mathbf{B}(t)}{dt}$.

We consider the simple cost function $J = \frac{1}{2}\epsilon^T \epsilon$, where $\epsilon = \bar{\mathbf{x}} - \hat{\bar{\mathbf{x}}}$, and hence for the estimation $\hat{\Lambda}$ we have $\hat{\Lambda} = \gamma(\bar{\mathbf{x}} - \hat{\bar{\mathbf{x}}})\bar{\mathbf{x}}^T$ (derivative of J with respect to $\bar{\mathbf{x}}$), where the estimation $\hat{\bar{\mathbf{x}}}(t+1) = e^{\hat{\Lambda}}\bar{\mathbf{x}}(t)$, and $\gamma > 0$ is chosen appropriately as the scaling gain. In optimization iterations, the estimated value of $\hat{\Lambda}$ at i^{th} iteration is as follows:

$$\hat{\Lambda}_i = \hat{\Lambda}_{i-1} + \hat{\Lambda}_{i-1}.$$

We use $\hat{\Lambda}_0 = \mathbf{I}_T \otimes (-\mathcal{L})$, the graph Laplacian of the explicit following-follower network (as initialization). The learned Λ may however, not be exactly structured as $\mathbf{I}_T \otimes (-\mathcal{L})$, due to dependence of topics in the state space, as well as the nonlinearity and non-homogeneity of the diffusion. The resulting error ϵ shall be considered for the estimation of the noise in the Kalman-Bucy filter as discussed next.

D. A Refined Prediction: Kalman-Bucy Filtering

In prediction applications, the actual states of some of the nodes are sometimes known, and we want to predict those of all remaining nodes. An example of this may be seen in social networks, where state of the hub nodes, such as famous people or users with less restrictive privacy policies are known to the public, and one is interested in predicting the state of other less accessible users. Having partial knowledge of the states of a fraction of the nodes in the network, changes the state prediction problem to a

Kalman predictor problem, and helps to refine the predicted states using a Kalman filter. We propose a Kalman-Bucy filter as the optimal linear predictor for our system, and we write the observation equation as $\bar{\mathbf{y}}(t) = (\mathbf{I}_T \otimes \mathbf{H})\bar{\mathbf{x}}(t) + \bar{\mathbf{v}}(t)$, with \mathbf{H} as a diagonal indicator matrix with 1 in all the observed entries, and 0 in all other entries. $\hat{\Lambda}$ having been learned (see above Section), Kalman-Bucy equations maybe written as:

$$\begin{aligned}\dot{\bar{\mathbf{x}}}(t) &= \hat{\Lambda}\bar{\mathbf{x}}(t) + \bar{\mathbf{w}}(t), \\ \bar{\mathbf{y}}(t) &= \mathcal{H}\bar{\mathbf{x}}(t) + \bar{\mathbf{v}}(t),\end{aligned}$$

where $\mathcal{H} = \mathbf{I}_T \otimes \mathbf{H}$ and the noises $\bar{\mathbf{w}}(t)$ and $\bar{\mathbf{v}}(t)$ are zero-mean white (temporally) processes, i.e., $E(\bar{\mathbf{w}}(t)\bar{\mathbf{w}}(s)^T) = \mathbf{Q}_t\delta(t-s)$, $E(\bar{\mathbf{v}}(t)\bar{\mathbf{v}}(s)^T) = \mathbf{R}_t\delta(t-s)$, $E(\bar{\mathbf{w}}(t)\bar{\mathbf{v}}(s)^T) = 0$. By considering small time intervals on discretization of the linear continuous time system ($\delta_t = 1$), one can write the state equation as $\bar{\mathbf{x}}(t+1) = \hat{\mathbf{F}}\bar{\mathbf{x}}(t) + \bar{\mathbf{w}}(t)$, where $\hat{\mathbf{F}} = \mathbf{I} + \hat{\Lambda}$:

$$\begin{aligned}\bar{\mathbf{x}}(t + \delta_t) &\simeq \bar{\mathbf{x}}(t) + \delta_t \dot{\bar{\mathbf{x}}}(t), \\ \bar{\mathbf{x}}(t + \delta_t) &\simeq \bar{\mathbf{x}}(t) + \delta_t (\hat{\Lambda}\bar{\mathbf{x}}(t) + \bar{\mathbf{w}}(t)), \\ \bar{\mathbf{x}}(t + \delta_t) &\simeq (\mathbf{I} + \delta_t \hat{\Lambda})\bar{\mathbf{x}}(t) + \delta_t \bar{\mathbf{w}}(t).\end{aligned}$$

Discretizing the Kalman-Bucy equations, gives us following discrete time equations:

$$\bar{\mathbf{x}}(t+1) = \hat{\mathbf{F}}\bar{\mathbf{x}}(t) + \bar{\mathbf{w}}(t), \quad (17)$$

$$\bar{\mathbf{y}}(t) = \mathcal{H}\bar{\mathbf{x}}(t) + \bar{\mathbf{v}}(t). \quad (18)$$

Having Eqns. (17 and 18) as the state and observation equations respectively, we can predict and refine the predicted states of the nodes using Algorithm (1):

Algorithm 1

Learning phase:

- 1: $\bar{\mathbf{x}}(t) \leftarrow \text{vec}(\mathbf{X}(t))$
- 2: $\hat{\Lambda} \leftarrow \mathbf{I}_T \otimes (-\mathcal{L})$ ▷ Initial state.
- 3: **repeat:**
- 4: $\bar{\mathbf{x}}(t+1) \leftarrow e^{\hat{\Lambda}}\bar{\mathbf{x}}(t)$
- 5: $\hat{\Lambda} \leftarrow \gamma(\bar{\mathbf{x}} - \hat{\bar{\mathbf{x}}})\bar{\mathbf{x}}^T$
- 6: $\hat{\Lambda} \leftarrow \hat{\Lambda} + \hat{\Lambda}$
- 7: **until** $\|\bar{\mathbf{x}} - \hat{\bar{\mathbf{x}}}\|_2 < \eta$. ▷ Convergence criteria.

Kalman filter prediction on test data:

- 1: $\mathbf{R}_{e,t} \leftarrow \mathbf{R}_t + \mathcal{H}\Pi_{t|t-1}\mathcal{H}^T$ ▷ Updating.
 - 2: $\hat{\mathbf{x}}_{t|t} \leftarrow \hat{\mathbf{x}}_{t|t-1} + \Pi_{t|t-1}\mathcal{H}^T\mathbf{R}_{e,t}^{-1}[\bar{\mathbf{y}}_t - \mathcal{H}\hat{\mathbf{x}}_{t|t-1}]$
 - 3: $\Pi_{t|t} \leftarrow \Pi_{t|t-1} - \Pi_{t|t-1}\mathcal{H}^T\mathbf{R}_{e,t}^{-1}\mathcal{H}\Pi_{t|t-1}$
 - 4: $\hat{\mathbf{F}} \leftarrow \mathbf{I} + \hat{\Lambda}$
 - 5: $\hat{\mathbf{x}}_{t+1|t} \leftarrow \hat{\mathbf{F}}\hat{\mathbf{x}}_{t|t}$ ▷ Predicting.
 - 6: $\Pi_{t+1|t} = \hat{\mathbf{F}}\Pi_{t|t}\hat{\mathbf{F}}^T + \mathbf{Q}_t$
-

The "learning phase" of the Algorithm (1) is estimating the supra-Laplacian matrix $\hat{\Lambda}$ (see above Section). The second phase of the algorithm, is refining the estimated state of the nodes. Note that \mathbf{R}_t is the covariance of the observational error, and $\hat{\mathbf{x}}_{t_2|t_1}$ denotes the linear prediction

of $\bar{\mathbf{x}}$ at time t_2 given observations up to and including time t_1 . The filter equation of a Kalman-Bucy filter [14], lines 1-3 of the algorithm, is given by:

$$\dot{\hat{\mathbf{x}}} = \hat{\Lambda}\hat{\mathbf{x}} + \mathbf{G}_t(\bar{\mathbf{y}}(t) - \mathcal{H}\hat{\mathbf{x}}(t)), \quad (19)$$

$$\mathbf{G}_t = \Pi_t\mathcal{H}^T\mathbf{R}_t^{-1}, \quad (20)$$

while the \mathbf{G}_t is the Kalman gain, and the state covariance Π_t satisfies the *Riccati* equation:

$$\dot{\Pi}_t = \hat{\Lambda}\Pi_t + \Pi_t\hat{\Lambda}^T + \mathbf{Q}_t - \mathbf{G}_t\mathbf{R}_t\mathbf{G}_t^T. \quad (21)$$

For simplicity, we further assumed that the errors in the state prediction and observation are Gaussian processes.

The designed algorithm shows the discrete time, state update of the Kalman predictor. The estimated states of the available nodes, $\mathcal{H}\hat{\mathbf{x}}(t)$, are compared with the state of the available nodes, $\bar{\mathbf{y}}(t)$, as measurements observed over time, to evaluate the extent of statistical noise and other inaccuracies in predicting phase. The Kalman gain \mathbf{G}_t , is tuned to assign accurate gain on the measurement (state of the available nodes) or follow the prediction model (estimating the state of the unknown nodes) more closely. More simply, the algorithm recursively learns from the error appeared in the estimation of the states of the available nodes and– by taking into account of the covariance of the measured error– refines the estimation of the states of the less accessible nodes.

E. Inter-layer Connectivity: Structural Robustness of an Interconnected Network

In an interconnected network setting, the inter-layer links play a crucial role in speed of diffusion between layers. Strong inter-layer links will cause a faster information diffusion among the layers and while, weak inter-layer connections will yield a set of independent layers [27]. In this section we use perturbation theory [23] to study the effect of weak inter-layer linkage on the connectivity of the over-all interconnected network. The second smallest eigenvalue of a Laplacian matrix (algebraic connectivity) helps us to uncover how close is the interconnected network is to break into multiple connected components [3, 17, 27]. To this end, we claim following proposition,

Proposition 2. *If an interconnected network has connected intra-layer networks but weak inter-layer links, the second smallest eigenvalue of the supra-Laplacian matrix of that network is equal to,*

$$\epsilon\mathbf{u}_n^T\mathcal{L}_I\mathbf{u}_n, \quad (22)$$

where \mathbf{u}_n is an eigenvector of the intra-layer supra-Laplacian matrix \mathcal{L}_L .

Proof. Similarly to Eqn. (5), write the supra-Laplacian matrix as a supra-Laplacian matrix of the intra-layer connectivity \mathcal{L}_L , and a supra-Laplacian matrix of a weak inter-layer $\epsilon\mathcal{L}_I$:

$$\mathcal{L} = \mathcal{L}_L + \epsilon\mathcal{L}_I, \quad (23)$$

where ϵ is a small positive number.

Denote \mathbf{v} and λ by respectively an eigenvector and the associated eigenvalue of the supra-Laplacian matrix \mathcal{L} , we have $\mathcal{L}\mathbf{v} = \lambda\mathbf{v}$. Similarly for the intra-layer supra-Laplacian, i.e., $\mathcal{L}_L\mathbf{u}_n = \lambda_n\mathbf{u}_n$, where $\{\mathbf{u}_n\}$ is an orthonormal set of vectors such that: $\mathbf{u}_n^T\mathbf{u}_m = \delta_{nm}$.

Since Laplacian matrices are symmetric and diagonalizable, an eigenvector of the supra-Laplacian matrix may be written as a linear combination of the eigenvectors of the intra-layer supra-Laplacian ($\mathbf{v} = \sum_m c_m\mathbf{u}_m$). We therefore, have:

$$\mathcal{L} \sum_m c_m \mathbf{u}_m = \lambda \sum_m c_m \mathbf{u}_m. \quad (24)$$

By right multiplying both hand-sides of the Eqn. (24) by \mathbf{u}_n^T and replacing the \mathcal{L} with Eqn. (23) we will readily have:

$$c_n \lambda_n + \epsilon \sum_m c_m \mathbf{u}_n^T \mathcal{L}_I \mathbf{u}_m = \lambda c_n. \quad (25)$$

Since we know that $\mathbf{v} = \mathbf{u}_n + \mathbf{o}(\epsilon)$, the eigenvector \mathbf{v} should be mainly in the same direction of one of the eigenvector's of the \mathcal{L}_L matrix with small perturbation. Therefore we can assume $c_m = o(\epsilon)$ and $c_n = o(1)$. In Eqn. (25), if we ignore the terms which have both ϵ and c_m multipliers we will get:

$$\lambda = \lambda_n + \epsilon \mathbf{u}_n^T \mathcal{L}_I \mathbf{u}_n. \quad (26)$$

Since \mathcal{L}_L is the Laplacian matrix of M independent intra-layers, it will have at least M unconnected components and M zero eigenvalues [27]. From Eqn. (26), we therefore can infer that if an interconnected network has connected intra-layer networks but weak inter-layer links, the second smallest eigenvalue of the supra-Laplacian matrix of that network is equal to $\epsilon \mathbf{u}_n^T \mathcal{L}_I \mathbf{u}_n$. \square

In section V we demonstrate through an experiment the effect of weak inter-layer connectivity of our state predictions model.

V. EXPERIMENTS

To evaluate and substantiate the theoretical interconnected-network model proposed above section, we conduct discrete-time simulations of our system in both "closed and open" system conditions.

A. Data Sets

Our experiments have been carried out using the following data sets,

- **Network of professors and publications** ($N \sim 100$): We assume a three-layer network, with two agent-layers for the professors (the first and the second layers) and one information-layer for publications (the third layer). In this data-set the agents are 79 professors at North Carolina State University, and the documents are 1000 abstracts of academic papers published by these professors 1990 to 2014. The agents in the first and the

second agent-layers are the same individuals with one-to-one connections between the agents. The first agent-layer is based on the number of papers two professors have co-authored in the same venue. $\mathbf{W}_{i,j}^{(1)}$ is the cumulative sum of papers published by professor i and j in the same venue. The second agent-layer reflects research group co-membership of two professors (with 8 different research groups considered).

$$\mathbf{W}_{i,j}^{(2)} = \begin{cases} 1 & \text{If professor } i \text{ and } j \text{ are in the same group,} \\ 0 & \text{Otherwise.} \end{cases} \quad (27)$$

The third layer (the information-layer) is based on the topical similarity of the produced documents, and is quantified by the inverse distance between the topical vectors ($T = 10$) of the documents. An ϵ -neighborhood criterion has been applied to break the weaker links,

$$\mathbf{W}_{i,j}^{(3)} = \begin{cases} \frac{1}{\|\mathbf{x}_i^{(3)} - \mathbf{x}_j^{(3)}\|_2} & \text{If larger than } \epsilon, \\ 0 & \text{Otherwise.} \end{cases} \quad (28)$$

$T = 10$ dimensional topic vectors of documents have been produced by performing the LDA topic modeling algorithm [1] on the raw text documents.

- **Network of Twitter users and Hashtags** ($N = 1000$): A two-layer network of Twitter users, (first layer) with ($N = 1000$) users and the similarity network between eight different popular Hashtags used by the Twitter users in June 2009. The Hashtags are as follows: #jobs, #spymaster, #neda, #140mafia, #tcot, #musicmonday, #Irancelection, #iremember. The agent-layer network is a directed graph, based on who is following whom on Twitter. The information-layer network is the similarity network between the Hashtags, where two Hashtags are considered similar if both have been mentioned in the same Tweet. We have used the Jaccard index [20] to build the network between the Hashtags, and the ϵ -neighborhood criterion has been applied to break the weaker links.
- $$\mathbf{W}_{m,n}^{(2)} = \begin{cases} \frac{|\text{Tweets which have both } m \text{ and } n|}{|\text{Tweets which have } m \text{ or } n|} & \text{If greater than } \epsilon. \\ 0 & \text{Otherwise.} \end{cases} \quad (29)$$
- **Network of Twitter users and Hashtags** ($N = 5000$): An exactly similar setting as the previously mentioned data set, with $N = 5000$ users.

B. Analysis of Results

Fig. (4) shows the results of our experiment on this data set (the network of professors and publications). The x-axis represents time, (and more specifically the consecutive years from Year 2000 until Year 2014). In our experiment, our goal is to predict the topical state of all the agents in each year by having the topical state of the agents in the previous year. More formally, we seek $\widehat{\mathbf{X}}^{(1)}(t)$ given $\widehat{\mathbf{X}}^{(1)}(t-1)$.

$$\text{Error measure}(t) = \frac{\|\widehat{\mathbf{X}}^{(1)}(t) - \mathbf{X}^{(1)}(t)\|_F}{\|\mathbf{X}^{(1)}(t)\|_F}, \quad (30)$$

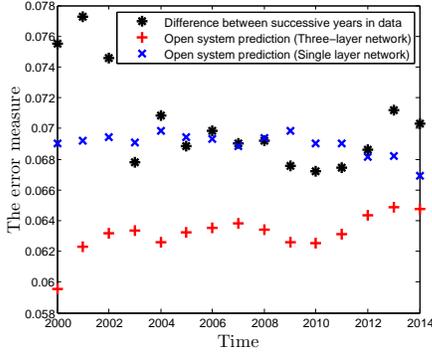


Fig. 4: Experiment for College Professor Network with 1000 publication documents.

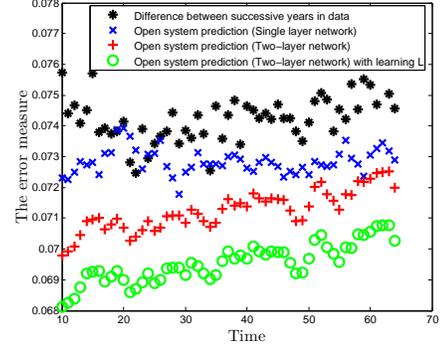


Fig. 5: Experiment for a Twitter network with 1000 agents and 8 Hashtags.

where $\mathbf{X}^{(1)}(t)$ is the ground truth matrix. The data over years 1990-1999 are used to learn the diffusion constants, which those from years 2000-2014 for testing. The graph in blue is the prediction errors by just considering a single-layer co-authorship network of agents/professors. The graph in red reflects the prediction error by considering a three-layer interconnected network of the heterogeneous nodes (as described in Section V-A). The graph in black shows the resulting changes in the topical states of the agents. This is hence a measure to reflect the difference in the yearly topical states of the agents, as in,

$$\frac{\|\mathbf{X}^{(1)}(t) - \mathbf{X}^{(1)}(t-1)\|_F}{\|\mathbf{X}^{(1)}(t)\|_F}. \quad (31)$$

Eqn. (31) is an upper bound of our prediction error, and any reasonable prediction method should have a smaller error than the one in Eqn. (31). We may indeed just assume no change in the agent states has occurred to achieve the error in Eqn. (31).

1) *Prediction using single layer network vs. heterogeneous network:* As may be seen in Fig. (4), the prediction method based on a three-layer network achieves a lower error than the prediction based on a single-layer network. Note, the single-layer network does not help in predicting the topical states of the agents. The reason is that there are only 79 agents in this experiment and the co-authorship network between the agents is not particularly suited to predict the future state of the agents. The three-layer network, on the other hand, considers the similarity between the documents present in the network, and accounts for more elaborate diffusion paths, thus enhancing the prediction phase. The average prediction improvement by the three-layer network over that of a single-layer network is 8 percent.

Figure 5 displays the results of our experiment on the second data set. The time (x-axis) reflects intervals of consecutive six-hour periods. In this experiment, recall, our goal is to predict the topical state of all agents at each time point when given the topical state of the agents at previous time points. The error metric is similar to that in

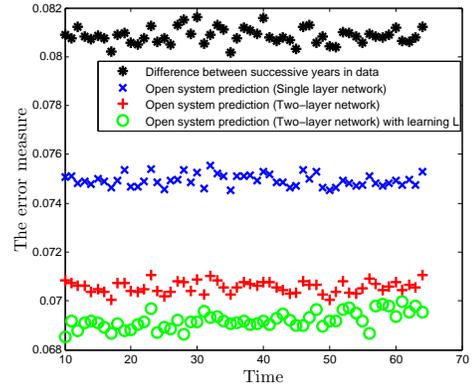


Fig. 6: Experiment over Twitter network with 5000 agents and 8 Hashtags.

Eqn. (30). The graph in blue reflects the prediction error by just considering a single-layer Twitter follower/following network between users. The graph in red displays the prediction error using a two-layer interconnected network of heterogeneous nodes (as described in Section V-A). The graph in black represents the changes in the topical states of users (Eqn. (31)). Finally, the graph in green shows the prediction error by first estimating the Laplacian matrix using the learning dataset and subsequently using it for the prediction phase. As may be seen in the figure, the prediction based on the two-layer network displays improvement over the prediction based on the single-layer network. We have used the data from time point 1 until time point 9 for learning the diffusion constants, and have used the data from Time-Point 10 to 64 for testing. The average prediction improvement using a two-layer network over a single-layer is 6 percent, while the prediction improvement using a single layer network over successive differences of data is 3 percent. We expect the improvement to further increase as the amount of data increases. Predicting the Laplacian matrix helps us on decreasing the prediction error. The average prediction improvement by this model over the single-layer method is about 10 percent. As in Fig. (4), the better prediction performance of the two-layer network is

due to the topical similarities of the documents. This may be attributed to the fact that this additional layer helps us discover certain topical interest of the users using similar Hashtags. For example, a Twitter user who has used a Hashtag #Iranelection, is probably going to use a Hashtag #neda (someone who died during the protests, after Iran’s presidential election in 2009) later. The lower prediction error by estimating the Laplacian matrix shows that the actual information diffusion structure can be different from the fixed connectivity imposed from the network.

Fig. (6) is the result of an experiment with similar conditions to that of Fig. (5), if not for the number of agents being 5000. As may be seen in the figure, the increased number of agents improves the prediction error of the two-layer network relative to the error upper bound. This effect is mainly due to two factors, (i) The increase in the number of agents in the network improve the information diffusion and the mixing. (ii) The increased network size provides a better estimate of the diffusion constants. The average prediction improvement achieved by the two-layer network is about 13 percent. The prediction by first estimating the Laplacian matrix, has about 15 percent improvement over the single layer prediction method.

In Fig. (7), we have used the term $\frac{\|\Sigma\|_F}{\|\mathbf{X}_0\|_F}$ (with $\mathbf{X}_0 = \mathbf{X}(0)$) to show the role of network size (number of nodes), on the extent to which the external sources affect the nodes. As the value of $\frac{\|\Sigma\|_F}{\|\mathbf{X}_0\|_F}$ term increases the second term on RHS of Eqn. (13) will have a higher weight in determining the state of the nodes and consequently, the states of the agents will change by higher variance. The figure suggests that as the network size increases the external influence of input decreases. This result can be attributed to expanding the system borders into external areas. In other word, as we increase the network size, we are including a larger number of factors which in turn, decreases the impact of auxiliary input.

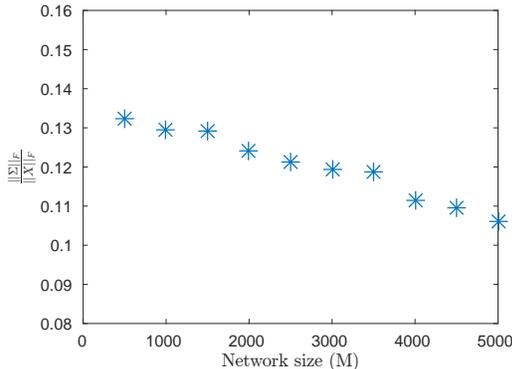


Fig. 7: Experiment over the Twitter network with different sizes.

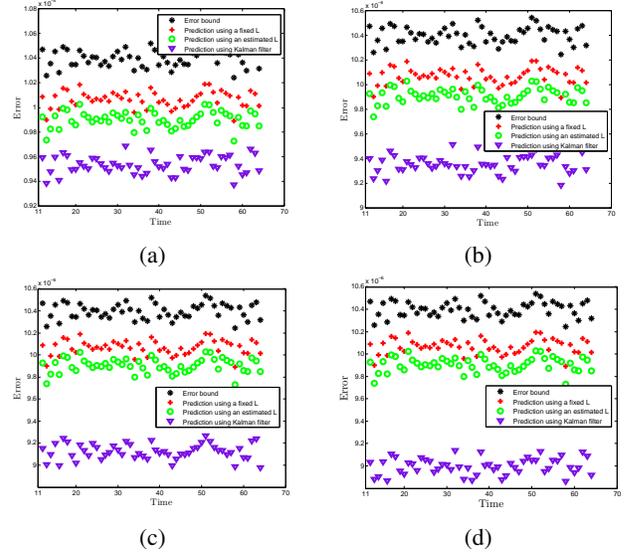


Fig. 8: Experiment over Twitter network with 300 agents. Predicting the state of the agents using a fixed Laplacian matrix, using an estimated Laplacian matrix, and Prediction using Kalman filter with 10 percent, 15 percent, 20 percent, and 25 percent of observation of state of all the agents in Figs. (a), (b), (c) and (d) respectively.

2) *Kalman Prediction Filtering*: Fig. (8) displays the results of an experiment in a small single layer dataset with 300 Twitter agents available. In this experiment, our goal is to predict the topical state of all agents at each time point, when given the topical state of agents at previous time points.

The black graphs in Fig. (8) show the resulting variation in the topical states of the agents overtime. This graph is an upper bound of our prediction error, and any reasonable prediction method should have a smaller error than this graph. The graphs in red reflect the prediction error by just considering a fixed Laplacian matrix constructed from a Twitter follower/following network between agents. The graphs in green display the prediction error by first learning the Laplacian matrix from the learning data set, and then using the estimated Laplacian matrix in the prediction. We have used the data from time-point 1 until time-point 10 for learning the diffusion structure (the Laplacian matrix), and have used the data from time-point 11 to 64 for testing. Finally, the graphs in purple represent the prediction error with a Kalman predictor, knowing a fraction of the states of agents a priori.

Figs. (8a), (8b), (8c) and (8d) are the same experiments with different observation sizes of 10 percent, 15 percent, 20 percent and 25 percent of state of all the agents in the network respectively. As may be seen in the figures, the prediction error based on the estimated Laplacian matrix yields a lower error than fully trusting the connectivity structure in the network. This as expected, is due to static connectivity network (usually demonstrates the physical or

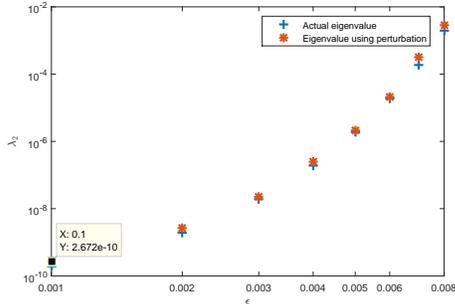


Fig. 9: Comparing the estimated value of the second smallest eigenvalue with the actual second smallest eigenvalue of the interconnected network.

online relation among the agents), falling short on effecting the actual underlying diffusion structure on the network. Having a prior partial knowledge of the agents enabled us to use a Kalman predictor to further refine our prediction. As we increase the size of the observation set, in Figs. (8b), (8c) and (8d) we get a smaller error in the prediction. This effect is due to the Kalman gain which refines the state of the system by the estimation error of the agent’s states.

3) Algebraic connectivity of an interconnected network:

Fig. (9) compares the actual second smallest eigenvalue (so called fiddler’s constant) of Laplacian matrix of an interconnected network with the estimated value using Eqn. (26). The x-axis shows different values of the ϵ , and the y-axis shows the second smallest eigenvalues of the supra-Laplacian matrix associated. As may be seen from the figure the estimated values closely match the actual values. Fig. (10) shows an experiment on the second dataset with different values of ϵ , and the effect of weak inter-layer connectivity on the state prediction error. As seen in Fig. (10), for small values of ϵ , the interconnected network is a single layer network and the prediction error is close to the single layer prediction error. However, as we increase the value of the ϵ , the inter-layer connectivity is getting stronger, and the prediction error decreases to the prediction error of a multi-layer network.

Based on Figs. (9, 10), one can use the Eqn. (26) to understand the strength of inter-layer connectivity in an interconnected network, and compute the algebraic connectivity. The algebraic connectivity in this case, helps us understand if the inter-layer connectivity is sequentially connecting all the intra-layers.

VI. CONCLUSION

In this work, we address the dynamics of an interconnected network by modeling changes in the topical states of agents using diffusion processes in the network. The intuition behind this model lies in the information diffusing on account of the possible interactions of the agents, which in turn may change their topical states over

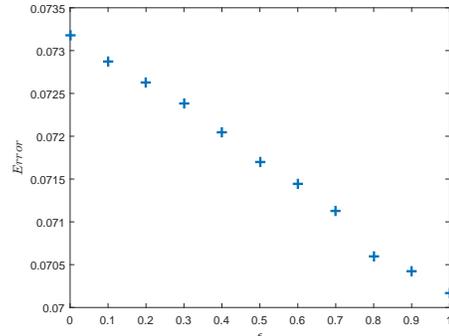


Fig. 10: The effect of weak inter-layer connectivity on the state prediction error.

time. To further generalize this idea, we propose to follow the different interactions of the agents in separate layers. We also consider the similarity of the data flowing in the network as another network layer that may help explain information diffusion in the interconnected network. To this end, we proposed a diffusion equation for the three-layer interconnected network. Moreover, we consider the external effect on each node by assuming that the whole system is a massive Brownian particle. In this model, the changes in the state of each node is a function of its interactions with other nodes and with the external effect modeled as Brownian motions. We test our method by experimenting on three real-world data sets. The results show that, the prediction for an interconnected network achieves a lower error. This is due to explicitly accounting for the data of the network as it also evolves among the agents. Also, we also showed that increasing the size of the network yields an improvement in the error.

REFERENCES

- [1] David M. Blei, Andrew Y. Ng, and Michael I. Jordan, *Latent dirichlet allocation*, *J. Mach. Learn. Res.* **3** (March 2003), 993–1022.
- [2] Charles D. Brummitt, Kyu-Min Lee, and K.-I. Goh, *Multiplexity-facilitated cascades in networks*, *Phys. Rev. E* **85** (2012Apr), 045102.
- [3] Fan RK Chung, *Spectral graph theory*, Vol. 92, American Mathematical Soc. 1997.
- [4] Faryad Darabi Sahneh and Caterina Scoglio, *Competitive epidemic spreading over arbitrary multilayer networks*, *Phys. Rev. E* **89** (2014Jun), 062817.
- [5] Scott C. Deerwester, Susan T Dumais, Thomas K. Landauer, George W. Furnas, and Richard A. Harshman, *Indexing by latent semantic analysis*, *JASIS* **41** (1990), no. 6, 391–407.
- [6] Mark Dickison, S. Havlin, and H. E. Stanley, *Epidemics on interconnected networks*, *Phys. Rev. E* **85** (2012Jun), 066109.
- [7] Francisco Escolano, Edwin R Hancock, and Miguel A Lozano, *Heat diffusion: Thermodynamic depth complexity of networks*, *Physical Review E* **85** (2012), no. 3, 036206.
- [8] Sebastian Funk and Vincent A. A. Jansen, *Interacting epidemics on overlay networks*, *Phys. Rev. E* **81** (2010Mar), 036118.
- [9] Sergio Gomez, Albert Diaz-Guilera, Jesus Gomez-Gardeñes, Conrad J Perez-Vicente, Yamir Moreno, and Alex Arenas, *Diffusion dynamics on multiplex networks*, *Physical review letters* **110** (2013), no. 2, 028701.
- [10] Manuel Gomez Rodriguez, Jure Leskovec, and Andreas Krause, *Inferring networks of diffusion and influence*, *Proceedings of the 16th acm sigkdd international conference on knowledge discovery and data mining*, 2010, pp. 1019–1028.

- [11] Manuel Gomez Rodriguez, Jure Leskovec, and Bernhard Schölkopf, *Structure and dynamics of information pathways in online media*, Proceedings of the sixth acm international conference on web search and data mining, 2013, pp. 23–32.
- [12] Herbert W. Hethcote, *The mathematics of infectious diseases*, SIAM review **42** (2000), no. 4, 599–653.
- [13] I.T. Jolliffe, *Introduction*, Principal component analysis, 1986, pp. 1–7.
- [14] Rudolph Emil Kalman, *A new approach to linear filtering and prediction problems*, Journal of Fluids Engineering **82** (1960), no. 1, 35–45.
- [15] Mikko Kivelä, Alexandre Arenas, Marc Barthelemy, James P Gleeson, Yamir Moreno, and Mason A Porter, *Multilayer networks*, arXiv preprint arXiv:1309.7233 (2013).
- [16] Byungjoon Min and K-I Goh, *Layer-crossing overhead and information spreading in multiplex social networks*, arXiv preprint arXiv:1307.2967 (2013).
- [17] Bojan Mohar, Y Alavi, G Chartrand, and OR Oellermann, *The laplacian spectrum of graphs*, Graph theory, combinatorics, and applications **2** (1991), 871–898.
- [18] Ingemar Näsell, *The quasi-stationary distribution of the closed endemic sis model*, Advances in Applied Probability (1996), 895–932.
- [19] Mark Newman, *Networks: an introduction*, Oxford University Press, 2010.
- [20] Tan Pang-Ning, Michael Steinbach, Vipin Kumar, et al. *Introduction to data mining*, Library of congress, 2006, pp. 74.
- [21] Dajun Qian, O. Yagan, Lei Yang, and Junshan Zhang, *Diffusion of real-time information in social-physical networks* (2012Dec), 2072–2077.
- [22] Manuel Gomez Rodriguez, David Balduzzi, and Bernhard Schölkopf, *Uncovering the temporal dynamics of diffusion networks*, arXiv preprint arXiv:1105.0697 (2011).
- [23] J. J. Sakurai and Jim J. Napolitano, *Modern quantum mechanics*, 2edition, Addison-Wesley.
- [24] Joaquin Sanz, Cheng-Yi Xia, Sandro Meloni, and Yamir Moreno, *Dynamics of interacting diseases*, Phys. Rev. X **4** (2014Oct), 041005.
- [25] Anna Saumell-Mendiola, M. Ángeles Serrano, and Marián Boguñá, *Epidemic spreading on interconnected networks*, Phys. Rev. E **86** (2012Aug), 026106.
- [26] Boris Shulgin, Lewi Stone, and Zvia Agur, *Pulse vaccination strategy in the sir epidemic model*, Bulletin of Mathematical Biology **60** (1998), no. 6, 1123–1148.
- [27] Albert Sole-Ribalta, Manlio De Domenico, Nikos E Kouvaris, Albert Diaz-Guilera, Sergio Gomez, and Alex Arenas, *Spectral properties of the laplacian of multiplex networks*, Physical Review E **88** (2013), no. 3, 032807.
- [28] Joshua B. Tenenbaum, Vin de Silva, and John C. Langford, *A global geometric framework for nonlinear dimensionality reduction*, Science **290** (2000), no. 5500, 2319–2323.
- [29] Warren S. Torgerson, *Multidimensional scaling: I. theory and method*, Psychometrika **17** (1952), no. 4, 401–419.
- [30] G. E. Uhlenbeck and L. S. Ornstein, *On the theory of the brownian motion*, Phys. Rev. **36** (1930Sep), 823–841.
- [31] Huijuan Wang, Qian Li, Gregorio D’Agostino, Shlomo Havlin, H. Eugene Stanley, and Piet Van Mieghem, *Effect of the interconnected network structure on the epidemic threshold*, Phys. Rev. E **88** (2013Aug), 022801.
- [32] Xuetao Wei, N.C. Valler, B.A. Prakash, I. Neamtiu, M. Faloutsos, and C. Faloutsos, *Competing memes propagation on networks: A network science perspective*, Selected Areas in Communications, IEEE Journal on **31** (2013June), no. 6, 1049–1060.
- [33] Osman Yağan and Virgil Gligor, *Analysis of complex contagions in random multiplex networks*, Phys. Rev. E **86** (2012Sep), 036103.
- [34] Dawei Zhao, Lixiang Li, Haipeng Peng, Qun Luo, and Yixian Yang, *Multiple routes transmitted epidemics on multiplex networks*, Physics Letters A **378** (2014), no. 10, 770–776.