Moment-Based Analysis of Spreading Processes from Network Structural Information

Victor M. Preciado* and Ali Jadbabaie†

Department of Electrical and Systems Engineering

University of Pennsylvania, Philadelphia, PA 19104‡

Abstract

The intricate structure of many large-scale networked systems has attracted the attention of the scientific community, leading to many results attempting to explain the relationship between network structural properties and dynamical performance. A common approach to study this relationship is the usage of synthetic network models in which the researcher can prescribe structural properties of interest, such as degree distributions. Researchers then estimate performance metrics of the synthetic network and study the effect of structural variations in these metrics. Although very common, this approach present a major flaw: Synthetic network models implicitly induce structural properties that are not directly controlled and can be relevant to the network dynamical performance. Therefore, it is difficult to isolate the role of a particular network property in the dynamical performance using synthetic networks. In this paper, we propose an alternative approach to overcome this flaw. Furthermore, our analysis unveils the set of structural properties that are most relevant to the network dynamical performance. We illustrate our approach by studying the dynamics of viral spreading processes in complex networks. Our analysis builds on algebraic graph theory and convex optimization to study how network structural properties constrain the behavior of viral spreading. We illustrate our approach with nontrivial numerical simulations in an online social network.

Keywords: Large-Scale Networks, Spreading Processes, Spectral Graph Theory, Convex Optimization

*Electronic address: preciado@seas.upenn.edu

†Electronic address: jadbabai@seas.upenn.edu

‡This work was supported by ONR MURI N000140810747.

I. INTRODUCTION

During the last decade, the complex structure of many real-world networked systems has attracted the attention of the scientific community [1]. The availability of massive databases describing these networks allow researchers to explore their structural properties with great detail. Statistical analysis of empirical data has unveiled the existence of multiple common patterns in a large variety of network properties, such as heavy-tailed degree distributions [2] or the small-world phenomenon [3]. Aiming to replicate these structural patterns, network researchers proposed a rich variety of synthetic network models. We can classify these models into two main families: *(i)* generative network models, and *(ii)* imitative network models. On the one hand, *generative* network models aim to explain the underlying mechanism that generates a particular network property. For example, the preferential attachment model, proposed by Barabási and Albert in [2], is a generative network model aiming to explain the presence of heavy-tailed degree distributions. On the other hand, the objective of *imitative* network models is not to understand the mechanism that originates a network property, but to build a synthetic network that imitates structural properties observed in real networks. Imitative networks are usually static random graphs ensembles, such as the classical Erdös-Rényi random graph and its generalizations [4]-[8].

Synthetic network models are useful to analyze the performance of dynamical processes taking place in realistic network topologies, as well as to predict and control network evolution. In this direction, a fundamental question is to understand the impact of a particular structural property in the network dynamical performance [9]. The most common approach to address this question is to use a synthetic network model in which one can prescribe the structural property under study. Most of the results in the literature are based on synthetic network models that incorporate structural features of apparent interest, such as degree distributions [4],[5], clustering [6], correlations [7], or hierarchy [8]. Researchers can then estimate performance metrics of the synthetic network and study the effect of structural variations in these metrics. Although this approach is very common in the literature, it presents two major flaws:

1. When a researcher uses a particular synthetic network model, he is making structural assumptions that go well beyond the property under study. Synthetic network models

implicitly induce structural properties that are not directly controlled and can be relevant to the network dynamical performance. For example, one can find synthetic network models with the same degree distribution but radically different dynamical performance [10]. The reason for this disparity in the performance is that while we fit a network property, e.g. the degree distribution, there are also many other network properties that are being implicitly modified. Since these properties are not directly controlled, it is difficult to isolate the role of a particular network property in the dynamical performance using synthetic network models.

2. There is no systematic technique to detect what structural properties may be relevant to the network dynamical performance. As a result, researchers choose structural properties to study driven by intuition, instead of analysis.

In this paper, we propose an alternative approach to study the role of structural network properties in the dynamical performance of a network. Our approach overcomes the two issues mentioned above, since it does not use synthetic models and also unveils what structural properties are most relevant to the network dynamical performance. In contrast with the synthetic network models, in our approach we do not estimate the values of network performance metrics, instead we derive bounds on these metrics imposed by the network structural properties. We illustrate our analysis by studying the dynamics of viral propagation in networks. In particular, we show how structural properties, such as the distribution of degrees, triangles and other substructures, constrain the rate of spreading of a virus in the network. Our results are relevant in many real scenarios, from disease spreading in social networks [11]-[14] to viral dissemination of information in computer networks [15],[16]. In our analysis, we exploit the close relationship between the spectral radius of the adjacency matrix of the network and the expected rate of spreading of a random viral infection [12]-[14]. Our work builds on algebraic graph theory and convex optimization to find optimal bounds on the spectral radius from network structural properties. We then use these bounds to study the relationships between network structural properties and the dynamical performance of spreading processes. Furthermore, our approach can also be extended to study the dynamics of other relevant network processes, such as synchronization of oscillators, consensus, and diffusion processes [17]-[22].

The rest of this paper is organized as follows. In the next section, we review graph-theoretical terminology and introduce definitions needed in our derivations. In Section III, we illustrate the relationship between spreading processes and spectral network properties using the susceptible-infected-susceptible model of viral infection. In Section IV, we derive explicit relationships between the moments of the eigenvalue spectrum and the distribution of certain network measurements. In Section V, we use convex optimization to compute optimal bounds on network spectral properties from its sequence of spectral moments. We illustrate our approach with nontrivial numerical simulations in online social networks in Section VI.

II. NOTATION

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ denote an undirected graph with n nodes, e edges, and no self-loops[1]. We denote by $\mathcal{V}(\mathcal{G}) = \{v_1, \ldots, v_n\}$ the set of nodes and by $\mathcal{E}(\mathcal{G}) \subseteq \mathcal{V}(\mathcal{G}) \times \mathcal{V}(\mathcal{G})$ the set of undirected edges of $\mathcal{G}$. If $\{v_i, v_j\} \in \mathcal{E}(\mathcal{G})$ we call nodes v_i and v_j *adjacent* (or neighbors), which we denote by $v_i \sim v_j$. We define a *walk* of length k from v_0 to v_k to be an ordered sequence of nodes $(v_0, v_1, ..., v_k)$ such that $v_i \sim v_{i+1}$ for $i = 0, 1, ..., k-1$. If $v_0 = v_k$, then the walk is closed. A closed walk with no repeated nodes (with the exception of the first and last nodes) is called a *cycle*. For example, *triangles*, *quadrangles* and *pentagons* are cycles of length three, four, and five, respectively.

Graphs can be algebraically represented via matrices. The adjacency matrix of an undirected graph $\mathcal{G}$, denoted by $A_\mathcal{G} = [a_{ij}]$, is an $n \times n$ symmetric matrix defined entry-wise as $a_{ij} = 1$ if nodes v_i and v_j are adjacent, and $a_{ij} = 0$ otherwise[2]. The eigenvalues of $A_\mathcal{G}$, denoted by $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$, play a key role in our paper. The spectral radius of $A_\mathcal{G}$, denoted by $\rho(A_\mathcal{G})$, is the maximum among the absolute values of its eigenvalues. We define the k-th spectral moment of the adjacency matrix $A_\mathcal{G}$ as

$$m_k(A_\mathcal{G}) = \frac{1}{n} \sum_{i=1}^{n} \lambda_i^k. \tag{1}$$

We define the set of neighbors of v as $\mathcal{N}_v = \{w \in \mathcal{V}(\mathcal{G}) : \{v, w\} \in \mathcal{E}(\mathcal{G})\}$. The number of neighbors of v is called the *degree* of node v, denoted by d_v. We can define several local

[1] An undirected graph with no self-loops is also called a *simple* graph.
[2] For simple graphs, $a_{ii} = 0$ for all i.

neighborhoods around a node v based on the concept of distance. Let $d(v,w)$ denote the *distance* between two nodes v and w (i.e., the minimum length of a walk from v to w). We say that v and w are k-hop neighbors if $d(v,w)=k$, and define the set of nodes within the k-th order neighbors of v as $\mathcal{N}_v^{(k)} = \{w \in \mathcal{V}(\mathcal{G}) : d(v,w) \leq k\}$. The set of nodes in $\mathcal{N}_v^{(k)}$ induces a subgraph $\mathcal{G}_v^{(k)} \subseteq \mathcal{G}$, with node-set $\mathcal{N}_v^{(k)}$ and edge-set $\mathcal{E}_v^{(k)}$ defined as the subset of edges connecting nodes in $\mathcal{N}_v^{(k)}$.

In some cases, it is also interesting to define another type of neighborhood based on the concepts of vertex and edge boundary. For a given subgraph $\mathcal{S} \subseteq \mathcal{G}$, we define its vertex boundary $\delta(\mathcal{S})$ and the edge boundary $\partial(\mathcal{S})$ as

$$\delta(\mathcal{S}) = \{i \in \mathcal{V} \backslash S \text{ s.t. } i \sim j, \text{ for } j \in S\},$$
$$\partial(\mathcal{S}) = \{\{i,j\} \in \mathcal{E} \text{ s.t. } i \in S,\, j \notin S\}.$$

We define the k-th order boundary neighborhood $\overline{\mathcal{G}}_v^{(k)}$ as the subgraph of $\mathcal{G}$ with node-set $\overline{\mathcal{N}}_v^{(k)} = \mathcal{N}_v^{(k)}$ and edge-set $\overline{\mathcal{E}}_v^{(k)} = \mathcal{E}_v^{(k-1)} \cup \partial\left(\mathcal{G}_v^{(k-1)}\right)$ (see Fig. 1). The distinction between $\mathcal{G}_v^{(k)}$ and $\overline{\mathcal{G}}_v^{(k)}$ is important in some real scenarios. For example, in online social networks, it is usually easy to discover who your friends' friends are (i.e., $\mathcal{N}_v^{(2)}$), as well as connections between your friends and your 2-hop friends. In contrast, it is usually more difficult to discover connections between two friends being 2 hops away from v. In this particular case, one could assume that each node has access to $\overline{\mathcal{G}}_v^{(2)}$ (but not to $\mathcal{G}_v^{(2)}$).

We say that a graphical property $\mathcal{P}_v$ of a node v is local with a radius r if $\mathcal{P}_v$ exclusively depends on the r-th neighborhood of v, i.e., $\mathcal{P}_v = f\left(\mathcal{N}_v^{(r)}\right)$. The most widely studied graphical properties are real-valued functions quantifying local structural properties, such as the degree of a node or the number of triangles in its neighborhood. In this paper, we study the relationship between the distribution of some real-valued local network measurements and the global behavior of spreading processes taking place in the network.

III. VIRUS SPREADING IN COMPLEX NETWORKS

In our work, we exploit the close relationship between spreading processes and spectral network properties. We illustrate this relationship using the standard Susceptible-Infected-Susceptible (SIS) model of viral spreading in a network [11]-[13]. (Similar results hold for other models of viral spreading.) In the SIS model, nodes represent individuals and edges

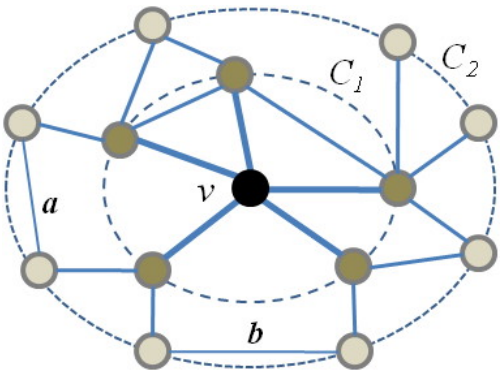

FIG. 1: The above figure represents the 2-hop neighborhood of a node v. The set $\mathcal{N}_v$ is the set of nodes lying on the circumference C_1, and $\mathcal{N}_v^{(2)}$ is the set of all nodes in the figure. The set $\delta\left(\mathcal{N}_v^{(1)}\right)$ is the set of nodes lying on the circumference C_2, and $\partial\left(\mathcal{N}_v^{(1)}\right)$ is the set of edges with ends in different circumferences. Notice that $\mathcal{G}_v^{(2)}$ is the whole graph in the figure, but $\overline{\mathcal{G}}_v^{(2)}$ does not include edges marked as a and b.

represent the structure of possible infections among them. Each individual can be in one of two possible states: susceptible or infected. A susceptible individual can be infected by an infected neighbor with a fixed probability β. If a node is infected, it stays infected during a (random) period of time during which it has the potential to infect its neighbors. In the following subsection, we study the relationship between the virus spreading and the spectral radius of the network adjacency matrix (more rigorous and thorough expositions can be found in [18] and [22]). In coming sections, we exploit this relationship to study the relationship between network structural properties and the dynamical performance of spreading processes.

A. Spreading Model and Spectral Results

Consider a population of n individuals interconnected via an undirected, labelled graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. Let $x_i[t] \in \{0, 1\}$ be a discrete-time variable that takes the value 1 if node i is infected at the time slot t, and 0 if node i is susceptible in that time slot. In each time slot, infected nodes attempt to contaminate their susceptible neighbors, where each infection attempt is successful with a given probability β, independent of other infection attempts. Hence, the probability of a susceptible node i to become infected in time slot t is

equal to $1-\prod_{j\in\mathcal{N}_i}\left(1-\beta x_j\left[t\right]\right)$. (This expression can be interpreted as the complementary probability that none of the infection attempts is successful.) Also, the probability of an infected node to recover in time slot t is equal to δ. The evolution of this stochastic process can be modeled using a Markov chain with 2^n states. According to the total probability theorem, the probabilities of infection evolve as follows:

$$\Pr\left(x_i\left[t+1\right]=1\right)=\left[1-\prod_{j\in\mathcal{N}_i}\left(1-\beta x_j\left[t\right]\right)\right]\Pr\left(x_i\left[t\right]=0\right) \tag{2}$$
$$+\left(1-\delta\right)\Pr\left(x_i\left[t\right]=1\right),$$

for $i=1,...,n$. Since $x_i\left[t\right]$ is a binary variable, we have that $\Pr\left(x_i\left[t\right]=1\right)=\mathbb{E}x_i\left[t\right]$. From a mean-field approximation [22], we have that:

$$z_i\left[t\right]=\left[1-\prod_{j\in\mathcal{N}_i}\left(1-\beta z_j\left[t\right]\right)\right]\left(1-z_i\left[t\right]\right)+\left(1-\delta\right)z_i\left[t\right], \tag{3}$$

where $z_i\left[t\right]=\mathbb{E}x_i\left[t\right]$.

Although different criteria of epidemic behavior can be found in the literature (see [14] for a recent review), we focus our attention in a criterion based on the local stability of the disease-free equilibrium [12]. Following this criterion, we linearize (3) around the disease-free equilibrium, $z_i\left[t\right]=0$ for all i, to obtain the following linear system of equations:

$$z_i\left[t\right]=\sum_{j\in\mathcal{N}_i}\beta z_j\left[t\right]+\left(1-\delta\right)z_i\left[t\right],$$

which can be written in matrix form as $\mathbf{z}\left[t\right]=\left[\beta A_{\mathcal{G}}+\left(1-\delta\right)I\right]\ \mathbf{z}\left[t\right]$. Hence, the disease-free equilibrium is locally stable if and only if the spectral radius of the transition matrix satisfies $\rho\left(\beta A_{\mathcal{G}}+\left(1-\delta\right)I\right)<1$, or equivalently

$$\rho\left(A_{\mathcal{G}}\right)<\frac{\delta}{\beta}. \tag{4}$$

In other words, Condition (4) determines a threshold for the disease-free equilibrium to become locally unstable, which is a necessary condition for an initial infection to become an epidemic.

We now present numerical simulations that illustrate the close connection between the spectral radius $\rho\left(A_{\mathcal{G}}\right)$ and the spreading of a virus in a real social network. In our simulations, we study the spreading of a SIS viral infection in a subgraph of $2,404$ nodes from an online social network.[3] We simulate the spreading process for values of δ/β above and

[3] We shall describe this dataset more thoroughly in Section VI.

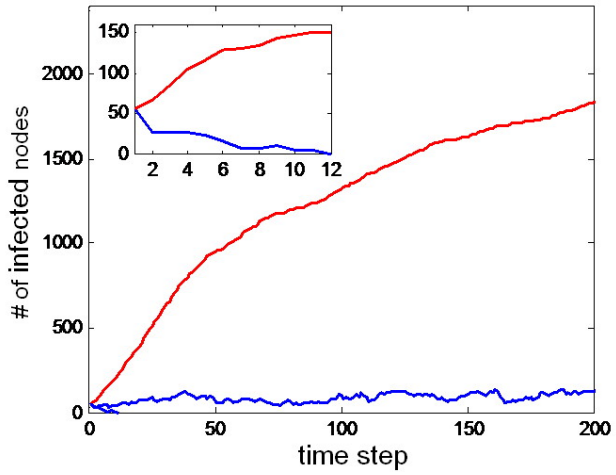

FIG. 2: Evolution of the number of nodes that are infected at a given time (blue) and the total number of nodes that are eventually infected (red), for $\delta/\beta = 70$ (in top-left subfigure) and $\delta/\beta = 50$ (main figure).

below the spectral radius of its adjacency matrix, $\rho(A_{\mathcal{G}}) = 60.90$, for a random initial infection with 2% probability. In Fig. 2, we plot the evolution of the total number of nodes that are eventually infected and the number of nodes that are infected at a given time, for two different values of δ/β. In the first simulation, we use $\delta = 0.70$ and $\beta = 0.010$, thus, $\delta/\beta = 70 > \rho(A_{\mathcal{G}})$ and the disease-free equilibrium is stable. We observe in Fig. 2 (top-left subfigure) how the initial viral infection quickly dies out in only 12 steps, and the total number of nodes that are eventually infected during the simulation is *148*, which is a small portion of the entire population. In the second simulation, we use the values $\delta = 0.50$ and $\beta = 0.010$, thus, $\delta/\beta = 50 < \rho(A_{\mathcal{G}})$. As we observe in Fig. 2, not only the disease-free equilibrium become unstable, but the infection persists over the entire time of the simulation, eventually infecting around $1,800$ nodes.

In Fig. 3, we illustrate the transition from quick death of an initial infection towards a global epidemic as we decrease the value δ/β. In particular, we fix the value $\beta = 0.010$, vary the value of δ in the interval [0.40:0.025:0.80], and run 50 simulations for each value of δ with different random initial infections (with a 2% probability of initial infection). In Fig. 3, we plot the average number of individuals that eventually become infected during the time of the simulation, as well as the associated typical deviation, for the 50 trials in each value of δ. We observe how, for values of $\delta/\beta > \rho(A_{\mathcal{G}}) = 60.90$, random initial infections tend to die out very quickly, infecting a small number of individuals. As we decrease the value of δ/β,

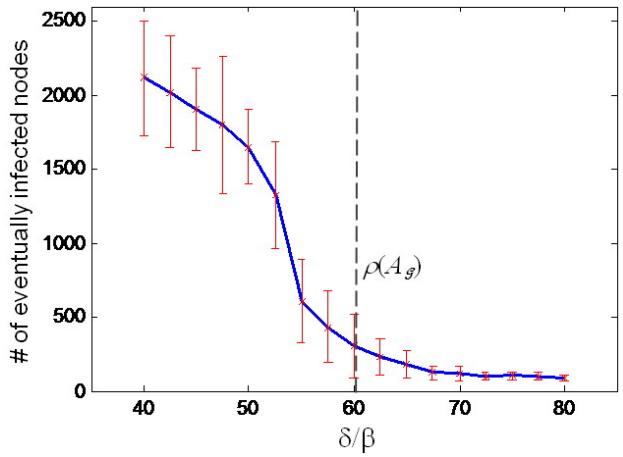

FIG. 3: Transition curve from global epidemic to quick death of an initial infection as we vary δ/β in [40:2.5:80]. For each value of δ/β, we plot the average and typical deviation of the total number of individuals that become eventually infected for 50 realizations of spreading dynamics with different random initial infections.

the average number of eventually infected nodes sharply increases, becoming very close to the total number of nodes when we move below a certain threshold (in our case, if we move δ/β below 53, more than half the population become eventually infected, in average).

Numerical analysis of the SIS model clearly indicates that for $\rho(A_{\mathcal{G}}) < \delta/\beta$ infections tend to persists over time and a significant fraction of the individuals in the network become eventually infected over time (see also [11]-[14] and references there in). This empirical observation is an apparent paradox, since the Markov chain describing the virus dynamics presents a unique absorbing state: the disease-free equilibrium (any other state is a transient state). Hence, from standard Markov chain theory [23], the infection should always dies out in finite time, which contradict the existence of infections persisting over time. Ganesh et al. [12] clarified this apparent contradiction by theoretically studying the behavior of the time T for an infection to die out. In particular, they found that Condition (4) is sufficient for $\mathbb{E}[T] = O(\log n)$ (which corresponds to a quick die out of the epidemic). They also derived a sufficient condition for $\mathbb{E}[T] = \Omega\left(e^{n^a}\right)$, for some $a > 0$. In this latter case, the time for an infection to die out is so large that it is usually mistaken by a persistent infection in numerical simulations.

In this subsection, we have illustrated the close connection between the spreading of a virus and the spectral radius $\rho(A_{\mathcal{G}})$. Similar results can be found for SIR infection models

in [12], for continuous-time models in [22], and for more general spreading processes in networked Markov chains in [18]. As a conclusion, the spectral radius of the adjacency matrix of a network is an efficient measure of its ability to spread virus/information.

B. Bounds on the Spectral Radius

In this subsection, we review some interesting results relating the distribution of network metrics with the spectral radius of the network. A common approach to study the relationship between structural properties and the spectral radius of a network is to use estimators based on random graph models [4]-[6]. For example, a popular estimator for the spectral radius in the complex networks literature is $\sum d_i^2 / \sum d_i$. This estimator corresponds to the expected spectral radius of an uncorrelated random networks[4] with degree sequence $\{d_i\}_{i=1}^n$. Since real-world networks are far from uncorrelated, these estimators are difficult to justify in most real applications.

In this paper, we propose an alternative approach that does not make any assumption on the network global structure. This approach consists in studying how a set of network measurements constrain some global network properties. In particular, we shall study how the distribution of network measurements impose bounds on the spectral radius of the adjacency matrix. We can find in the literature some bounds on the spectral radius related to our work. For a graph $\mathcal{G}$, with n nodes and e edges, we have the following upper bounds for the spectral radius:

$$\begin{aligned} u_1 &= \sqrt{2e - (n-1)\, d_{\min} + (d_{\min} - 1)\, d_{\max}}, \text{ in [24]}, \\ u_2 &= \max\left\{ \sqrt{d_i m_j},\ (i,j) \in E \right\}, \text{ in [25]}, \\ u_3 &= \min_{1 \le i \le n} \frac{d_i - 1 + \sqrt{(d_i+1)^2 + 4\,(i-1)\,(d_{\max} - d_i)}}{2}, \text{ in [26]}, \end{aligned} \tag{5}$$

where $d_{\min}$ and $d_{\max}$ are the minimum and maximum degrees of $\mathcal{G}$, and $m_i = \frac{1}{d_i} \sum_{j \in \mathcal{N}_i} d_j$. Notice that none of the above bounds take into account the distribution of triangles in the graph. Since the presence of triangles and other cycles has a strong influence on the

[4] In an uncorrelated random graph, the probability that two nodes with degrees d_1 and d_2 are connected is proportional to $d_1 d_2$.

behavior of spreading processes in graphs, we expect these bounds to perform poorly in real applications. In the following sections, we propose a methodology to derive lower bounds on the spectral radius taking into account not only the distribution of degrees, but also other pieces of structural information.

IV. MOMENT-BASED ANALYSIS OF THE ADJACENCY SPECTRUM

Algebraic graph theory provides us with tools to relate structural and spectral properties of a network. The following result relates the k-th spectral moment of $A_{\mathcal{G}}$ (a spectral property) with the number of closed walks of length k in $\mathcal{G}$ (a structural property) [27]:

Lemma IV.1 *Let $\mathcal{G}$ be a simple graph. The k-th spectral moment of the adjacency matrix of $\mathcal{G}$ can be written as*

$$m_k(A_{\mathcal{G}}) = \frac{1}{n}\left|\Psi_{\mathcal{G}}^{(k)}\right|, \tag{6}$$

where $\Psi_{\mathcal{G}}^{(k)}$ is the set of all closed walks of length k in $\mathcal{G}$ [5].

In the following subsections, we build on Lemma IV.1 to compute the spectral moments of a network from the distribution of certain structural measurements. In Section V, we shall use these spectral moments to bound the spectral radius of $A_{\mathcal{G}}$ via semidefinite programming.

A. Low-Order Spectral Moments

From (6), we can easily compute the first three moments of $A_{\mathcal{G}}$ in terms of structural properties as follows [27]:

Corollary IV.2 *Let $\mathcal{G}$ be a simple graph with adjacency matrix $A_{\mathcal{G}}$. Denote by d_i and t_i the number of edges and triangles touching node $i \in \mathcal{V}(\mathcal{G})$, respectively. Then,*

$$m_1(A_{\mathcal{G}}) = 0,\ m_2(A_{\mathcal{G}}) = \frac{1}{n}\sum_{i\in\mathcal{V}(\mathcal{G})} d_i,\ \text{and}\ m_3(A_{\mathcal{G}}) = \frac{1}{n}\sum_{i\in\mathcal{V}(\mathcal{G})} 2t_i. \tag{7}$$

[5] We denote by $|Z|$ the cardinality of a set Z.

Hence, the second and third moments can be computed via a simple aggregation of the number of edges and triangles touching each node in $\mathcal{G}$. These moments can be also expressed in terms of the total number of edges and triangles in $\mathcal{G}$, denoted by e and Δ, respectively. Since $e = \frac{1}{2}\sum_i d_i$ and $\Delta = \frac{1}{3}\sum_i t_i$ [27], we have that:

$$m_2(A_{\mathcal{G}}) = 2e/n, \text{ and } m_3(A_{\mathcal{G}}) = 6\,\Delta/n,$$

where the coefficients 2 and 6 in the above expressions can be interpreted as the number of closed walks enabled by the presence of an edge or a triangle, respectively. The computation of higher-order moments requires a more elaborated combinatorial analysis. In coming subsections, we show that as we increase the order of the spectral moments, more elaborated structural properties arise in our expressions. We include details for the fourth and fifth spectral moments in the following subsections.

B. Higher-Order Spectral Moments

In this subsection, we compute the fourth and fifth spectral moments based on structural information.

1. Fourth-Order Spectral Moment

A combinatorial analysis of (6) for $k = 4$ gives us the following result:

Lemma IV.3 *Let $\mathcal{G}$ be a simple graph with adjacency matrix $A_{\mathcal{G}}$. Denote by q_i and d_i the number of quadrangles and edges touching node $i \in \mathcal{V}(\mathcal{G})$, respectively. Then,*

$$m_4(A_{\mathcal{G}}) = \frac{1}{n}\left[\sum_{i\in\mathcal{V}(\mathcal{G})} 2q_i + 4\binom{d_i}{2} + d_i\right]. \tag{8}$$

Proof. We compute the fourth moment from (6) by counting the number of closed walks in $\mathcal{G}$. In Fig. 4, we enumerate the possible types of closed walks of length 4 in a simple graph. We can count the number of closed walks of each particular type in terms of local network metrics as follows:

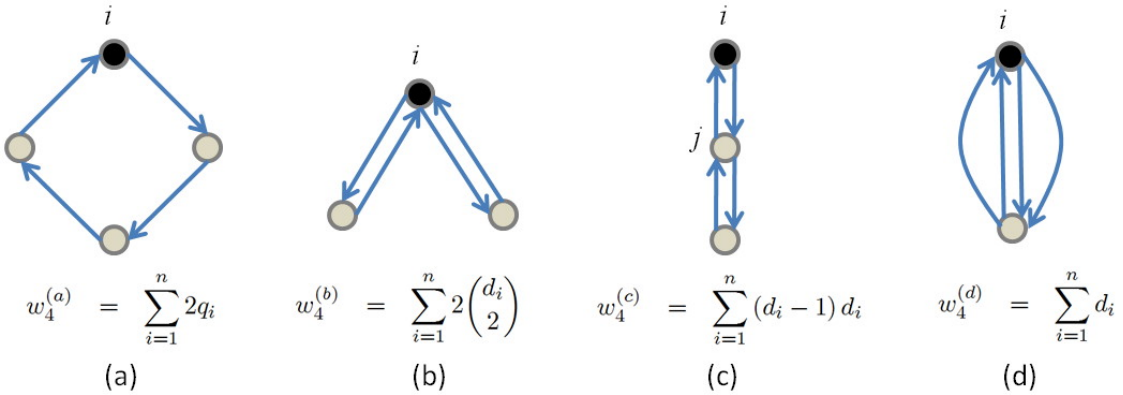

FIG. 4: Enumeration of the possible types of closed walks of length 4 in a graph with no self-loops. The classification is based on the structure of the subgraph underlying each closed walk. For each walk type, we also include an expression that corresponds to the number of closed walks of that particular type in terms of local network metrics.

Type (a) The number of closed walks in this type is equal to twice the number of quadrangles, where the coefficient 2 in $w_4^{(a)}$ accounts for the two possible directions, clockwise and counterclockwise, to walk each quadrangle.

Type (b) The number of this type of graphs, that we denote as wedge graphs, can be written in terms of the degrees. Notice that the number of wedge graphs centered at node i is equal to $\binom{d_i}{2}$, and that there are two possible directions one can choose for the first step in the walk. The expression for $w_4^{(b)}$ comes from summing over all possible starting points, $i = 1, ..., n$.

Type (c) The number of closed walks on this type of graphs, called 2-chain graphs, can also be written in terms of the degrees as:

$$w_4^{(c)} = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}\,(d_j - 1) = \sum_{j=1}^{n} (d_j - 1)\, d_j.$$

Type (d) The number of closed walks of this type starting at node i is equal to d_i, thus, $w_4^{(d)} = \sum_{i=1}^{n} d_i$.

Hence, we obtain (8) by summing up all the above contributions (and simple algebraic manipulations). ■

Lemma IV.3 provides an expression to compute the fourth spectral moment in terms of local metrics, i.e., the distribution of degrees and quadrangles. (Notice that d_i and q_i can

be extracted from $\mathcal{G}_i^{(1)}$ and $\overline{\mathcal{G}}_i^{(2)}$, respectively.) Hence, $m_4\left(A_{\mathcal{G}}\right)$ can be computed as a simple aggregation of local network properties with radius 2. We illustrate Lemma IV.3 in the following example.

Example IV.1 *Consider the n-ring graph, R_n (without self-loops). The eigenvalues of the adjacency matrix of the ring graph are $\lambda_i\left(A_{R_n}\right) = 2\cos i\frac{2\pi}{n}$, for $i = 0, 1, ..., n-1$. Hence, the 4-th moment is equal to $m_4(A_{R_n}) = \frac{1}{n}\sum_{i=0}^{n-2}\left(2\cos i\frac{2\pi}{n}\right)^4$, which (after some computations) can be found to be equal to 6 for $n \notin \{2, 4\}$. We can reach this same result by directly applying (9), without performing an eigenvalue decomposition, as follows. In the ring graph, we have that $d_i = 2$ and $q_i = 0$, for $n \notin \{2, 4\}$. Hence, from (8), we directly obtain $m_4(A_{R_n}) = 6$, $n \notin \{2, 4\}$.*

The fourth spectral moment can be rewritten in terms of aggregated quantities, such as the total number of quadrangles and edges, and the sum-of-squares of the degrees, as follows:

Corollary IV.4 *Let $\mathcal{G}$ be a simple graph. Denote by e and Q the total number of edges and quadrangle in $\mathcal{G}$, respectively, and define $W_2 = \sum_{i=1}^{n} d_i^2$. Then,*

$$m_4\left(A_{\mathcal{G}}\right) = \frac{1}{n}\left[8Q + 2W_2 - e\right]. \tag{9}$$

Proof. The proof comes straightforward from (8) by substituting $\sum_{i=1}^{n} q_i = 4Q$ and $\sum_{i=1}^{n} d_i = 2e$. ■

Hence, we do not need to have access to the detailed distribution of quadrangles and degrees in $\mathcal{G}$ to compute the 4-th moment, we only need to know the aggregated quantities, Q, W_2, and e, that appear in (9).

2. Fifth-Order Moment

Lemma IV.5 *Let $\mathcal{G}$ be a simple graph. Denote by p_i, t_i, and d_i the number of pentagons, triangles, and edges touching node $i \in \mathcal{V}\left(\mathcal{G}\right)$, respectively. Then,*

$$m_5\left(A_{\mathcal{G}}\right) = \frac{1}{n}\left[\sum_{i\in\mathcal{V}(\mathcal{G})} 2p_i + 10t_i d_i - 10t_i\right]. \tag{10}$$

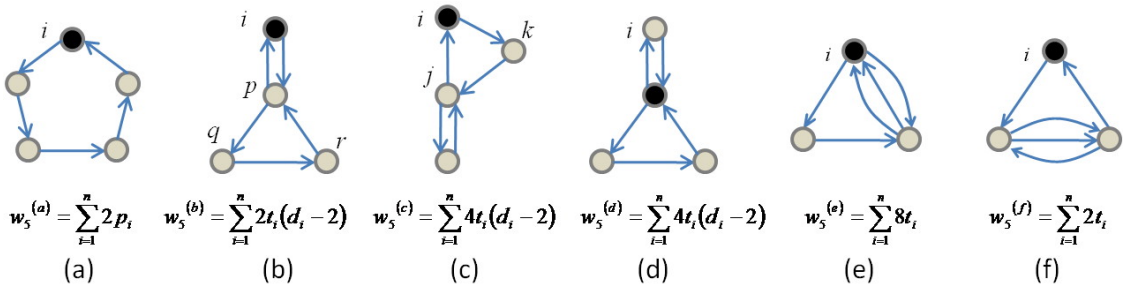

FIG. 5: Possible types of closed walks of length 5 in a simple graph. The classification is based on the structure of the subgraph underlying the closed walks. For each walk type, we also include an expression that corresponds to the number of closed walks of that particular type in terms of local network metrics.

Proof. The proof follows the same structure as that of Lemma IV.3. A graphical representation of the types of closed walks of length 5 is provided in Fig. 5. Details regarding the counting of closed walks of each particular type can be found in Appendix I. ■

Remark IV.1 *Notice that d_i and t_i can be extracted from $\mathcal{G}_i^{(1)}$, and q_i can be extracted from $\mathcal{G}_i^{(2)}$, respectively. Therefore, $m_5(A_\mathcal{G})$ can be computed from a simple aggregation of local properties with radius 2. An efficient strategy to distributedly aggregate the quantities involved in the first five spectral moments is to follow these steps:*

(Step 1) Each node locally extracts d_i, t_i, q_i, and p_i from $\mathcal{G}_i^{(2)}$.

(Step 2) Each node locally computes the expressions inside the summations of (7), (8) and (10).

(Step 3) Run a distributed averaging algorithm in the network to compute the average of the quantities in the brackets.

Lemma (IV.5) expresses the fifth spectral moment of $A_\mathcal{G}$ in terms of local network metrics. We can rewrite (10) in terms of aggregated quantities as follows:

Corollary IV.6 *Let $\mathcal{G}$ be a simple graph. Denote by Δ and Π the total number of triangles and pentagons in $\mathcal{G}$, respectively. Define the degree-triangle correlation as $\mathcal{C}_{dt} = \sum_i d_i t_i$. Then,*

$$m_5(A_\mathcal{G}) = \frac{1}{n}\left[10\Pi + 10\mathcal{C}_{dt} - 30\Delta\right]. \tag{11}$$

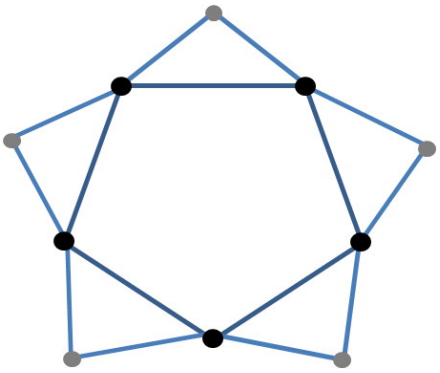

FIG. 6: Graph considered in Example IV.2.

Proof. The proof comes from (10) taking into account that $\sum_{i=1}^{n} p_i = 5\Pi$ and $\sum_{i=1}^{n} t_i = 3\Delta$. ■

The above Corollary indicates that there is a direct relationship between the degree-triangle correlation, $\mathcal{C}_{dt}$, of a graph and its spectral properties. Although nontrivial variations of this correlation term have been reported for real-world networks [28], little attention has been paid to it in the literature. Corollary IV.6 provides a tool to analyze the effect of $\mathcal{C}_{dt}$ on the behavior of spreading processes in the network (via the spectral radius). We illustrate the usage of Corollary IV.6 in the following example:

Example IV.2 *Consider the graph in Fig. 6. In this example, we use (10) to directly compute the fifth spectral moment of this graph without performing an explicit eigenvalue decomposition. In this case, we have that* $\Pi = 1$ *and* $\Delta = 5$. *Since* $d_i t_i = 8$ *for nodes in the pentagon, and* $d_j t_j = 2$ *for nodes outside the pentagon, we have that* $\mathcal{C}_{dt} = 50$. *Hence, the fifth moment is* $m_5(A_{\mathcal{G}}) = \frac{1}{10}\left[10 - 150 + 500\right] = 36$ *(which coincides with the value obtained from an explicit eigenvalue decomposition).*

The main advantage of our results may not be apparent in networks with simple, regular structure. For these networks, an explicit eigenvalue decomposition is usually easy to compute and there may be no need to look for alternative ways to compute spectral properties. On the other hand, in the case of large-scale complex networks, the structure of the network can be very intricate —in many cases not even known exactly— and an explicit eigenvalue decomposition can be very challenging to compute, if not impossible. It is in these cases when the alternative approach proposed in this paper is most useful. In the

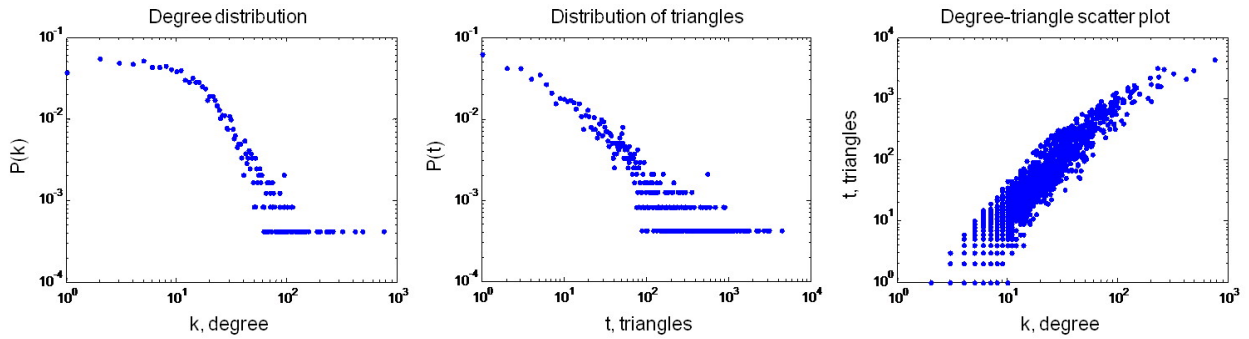

FIG. 7: In the left and center figures, we plot the distributions of degrees and triangles in log-log scale, of the social network under study, respectively. In the right figure, we include a scatter plot where the coordinates of each point corresponds to the number of triangles touching a node and its degree in log-log scale.

following subsection, we illustrate our approach with empirical data from an online social network.

C. Empirical Results for an Online Social Network

The real network under study is a subgraph of Facebook with $2,404$ nodes and $22,786$ edges obtained from crawling the graph in a breadth-first search around a particular node (the dataset can be found in [29]). Although the approach proposed in this section is meant to be used for much larger networks, we illustrate our results with this medium-size graph in order to compare our results with the results obtained from an explicit eigenvalue decomposition of the complete network topology.

In this example, we assume that each node i has access to its second-order neighborhood, $\mathcal{G}_i^{(2)}$. Hence, each node can compute its degree d_i, as well as the number of triangles t_i, quadrangles q_i, and pentagons p_i touching it. In Fig. 7, we show the degree and triangle distributions, as well as a scatter plot where the coordinates of each point corresponds to the number of triangles touching a node and its degree.

We can aggregate, via simple averaging, those quantities that are relevant to compute

the spectral moments. We obtain:

$$\begin{aligned} e/n &= \tfrac{1}{2n}\sum d_i = 9.478, & \Delta/n &= \tfrac{1}{3n}\sum t_i = 28.15, \\ Q/n &= \tfrac{1}{4n}\sum q_i = 825.3, & \Pi/n &= \tfrac{1}{5n}\sum p_i = 31,794, \\ W_2/n &= \tfrac{1}{n}\sum d_i^2 = 1,318, & \mathcal{C}_{dt}/n &= \tfrac{1}{n}\sum d_i t_i = 8,520. \end{aligned}$$

Hence, from expressions (7), (9), and (11), we obtain the following values for the spectral moments:

$$\begin{aligned} &m_1(A_{\mathcal{G}}) = 0, && m_2(A_{\mathcal{G}}) = 18.95, && m_3(A_{\mathcal{G}}) = 168.90, \\ &m_4(A_{\mathcal{G}}) = 9,230, && m_5(A_{\mathcal{G}}) = 402,310. \end{aligned} \tag{12}$$

We close this section with several concluding remarks. First, although one could derive closed expressions for moments of order greater than 5, the terms involved could not be computed from the structural information extracted from 2-hop neighborhoods. For example, in order to compute the sixth moment, a node should be able to count the number of hexagons touching it (closed walks of length 6). This quantity cannot be locally computed if the node only has access to its 2-hop neighborhood.[6] As a general rule, *if nodes have access to their r-hop neighborhoods, we can compute the sequence of spectral moments up to order $k_{\max} = 2r + 1$ by aggregating local structural information* [30]. In real-world networks, the average size of a *3*-hop neighborhood is usually very large. (This is a consequence of the 'Six Degrees of Separation' phenomenon, since 3-hop neighborhoods have diameter 6.) Hence, the computational cost of counting subgraphs in 3-hop neighborhoods become extremely costly.

In this section, we have derived expressions to compute global spectral information, i.e., the first five spectral moment of $A_{\mathcal{G}}$, from the distribution of local structural properties in the network. In the next section, we use semidefinite programming to extract bounds on spectral properties of interest from a sequence of spectral moments. In Section VI, we shall present implications of our results in the behavior of spreading processes in networks.

[6] Although other alternatives could be proposed to surpass this issue, such as a message-passing approach, we consider only structural metrics that can be *directly* extracted from the 2-hop neighborhoods.

V. OPTIMAL BOUNDS FROM SPECTRAL MOMENTS

In this section, we introduce an approach to derive optimal bounds on the spectral radius of $A_{\mathcal{G}}$ from a sequence of spectral moments. Since we have expressions for the spectral moments in terms of structural properties, these bounds relate the spectral radius of a network with its structural properties. Since there are already many efficient algorithms to compute the spectral radius of large networks, it is worth mentioning that our objective is not to compute the spectral radius, but to understand the role of network structural properties in the spectral radius and the network's ability to spread information virally. For this purpose, we adapt a technique proposed in [31] and [32] to derive optimal probabilistic bounds of random variables with a given sequence of moments. In order to use this technique, we introduce a probabilistic interpretation of a network spectrum and its spectral moments. For a simple graph $\mathcal{G}$, we define its spectral density as,

$$\mu_{\mathcal{G}}(x) = \frac{1}{n}\sum_{i=1}^{n}\delta(x-\lambda_i), \tag{13}$$

where $\{\lambda_i\}_{i=1}^{n}$ are the eigenvalues of the adjacency matrix $A_{\mathcal{G}}$. Notice that the moments of a random variable $X \sim \mu_{\mathcal{G}}$ are equal to the spectral moments of $A_{\mathcal{G}}$, i.e.,

$$\begin{aligned}\mathbb{E}_{\mu_{\mathcal{G}}}\left(X^k\right) &= \int x^k \mu_{\mathcal{G}}(x)\,dx \\ &= \frac{1}{n}\sum_{i=1}^{n}\int x^k \delta(x-\lambda_i)\,dx \\ &= \frac{1}{n}\sum_{i=1}^{n}\lambda_i^k = m_k(A_{\mathcal{G}}),\end{aligned}$$

for all $k \geq 0$. Furthermore, for a given Borel measurable set T, we have

$$\Pr(X \in T) = \int_{x\in T}\mu_{\mathcal{G}}(x)\,dx = \frac{1}{n}\left|\{\lambda_i : \lambda_i \in T\}\right|.$$

In other words, the probability of the random variable $X \sim \mu_{\mathcal{G}}$ being in a set T is proportional to the number of eigenvalues of $A_{\mathcal{G}}$ in T.

As mentioned before, we assume that we are studying very large complex networks and that we do not have access to the complete network topology. Instead, each node in the network can only access a local neighborhood around it. In this context, it is not possible to explicitly compute the set of eigenvalues $\{\lambda_i\}_{i=1}^{n}$. On the other hand, in Section IV, we

showed how to compute a sequence of $2r+1$ spectral moments from structural information measured in r-hop neighborhoods. In this section, we show how to extract global spectral information from this sequence of spectral moments.

We illustrate the idea behind our approach with the following simple example. Assume the extreme case in which we only have access to the degrees of the network, i.e., $\{d_i\}_{i=1}^n$. From Corollary IV.2, we have that the first and second spectral moments of the adjacency are $m_1 = 0$ and $m_2 = \sum_{i=1}^n d_i/n$. From these moments, we have the following result:

Claim V.1 *For a simple graph $\mathcal{G}$ with n nodes and degree sequence $\{d_i\}_{i=1}^n$, the spectral radius of its adjacency matrix $A_\mathcal{G}$ satisfies*

$$\rho\left(A_\mathcal{G}\right) \geq \sqrt{\frac{\sum_{i=1}^n d_i}{n}}. \tag{14}$$

Proof. Consider a random variable $X \sim \mu_\mathcal{G}$, as defined in (13). From Corollary IV.2, we know that $\mathbb{E}X = m_1 = 0$ and $\mathbb{E}X^2 = m_2 = \sum_{i=1}^n d_i/n$. We also have that

$$\begin{aligned}\Pr\left(|X| \geq \sqrt{m_2}\right) &= \int_{|x|\geq\sqrt{m_2}} d\mu_\mathcal{G}\left(x\right) \\ &= \frac{1}{n}\left|\{\lambda_i : |\lambda_i| \geq \sqrt{m_2}\}\right|. \end{aligned} \tag{15}$$

Then, by contradiction, we can prove that $\Pr\left(|X| \geq \sqrt{m_2}\right) > 0$, as follows. If we had that $\Pr\left(|X| \geq \sqrt{m_2}\right) = 0$, then, from (15), we would have that $|\lambda_i| < \sqrt{m_2}$ for all i. This would imply that

$$m_2 = \int_\mathbb{R} x^2 d\mu_\mathcal{G}\left(x\right) = \frac{1}{n}\sum_{i=1}^n \lambda_i^2 < m_2,$$

which is a contradiction. Therefore, since $\Pr\left(|X| \geq \sqrt{m_2}\right) = \frac{1}{n}\left|\left\{\lambda_i : |\lambda_i| \geq \sqrt{m_2}\right\}\right| > 0$, we have that $\rho\left(A_\mathcal{G}\right) = \max\lambda_i \geq \sqrt{m_2} = \sqrt{\sum_{i=1}^n d_i/n}$. ■

Hence, using the first two spectral moments we can derive a lower bound on the spectral radius from the degree sequence of the network. In what follows, we generalize this idea to more sophisticated local network metrics and higher-order spectral moments.

A. Optimal Bounds from Higher-Order Moments

In our derivations, we pay special attention to the following problem:

Problem 1 (Moment Problem) Given a sequence of moments $(m_0, ..., m_k)$, and Borel measurable sets $T \subseteq \Omega \subseteq \mathbb{R}$, we are interested in computing:

$$\begin{aligned} Z_P = \ & \max_\mu \ \int_T 1 \, d\mu \\ & s.t. \ \int_\Omega x^j \, d\mu = m_j, \ \textit{for } j = 0, 1, ..., k. \end{aligned} \tag{16}$$

where $\mu \in \mathbb{M}(\Omega)$, $\mathbb{M}(\Omega)$ being the set of positive Borel measures supported by Ω.

In other words, for a random variable $X \sim \mu$, we are interested in computing the maximum value of $\int_T 1 \, d\mu = \Pr(X \in T)$, when a sequence of moments of X is known. The solution to this problem is an extension to the classical Markov and Chebyshev's inequalities in probability theory when moments of order greater than 2 are available. In [31] and [32], it was shown that the optimal value of Z_P can be efficiently computed by solving a single semidefinite program using a dual formulation. Before we introduce this dual formulation, it is convenient to discuss some details regarding the feasibility of this problem.

A sequence of moments $\mathbf{m}_k = (m_0, m_1, ..., m_k)$ is said to be feasible in Ω if there exists a measure $\mu \in \mathbb{M}(\Omega)$ whose moments match those in the sequence $\mathbf{m}_k$. In general, an arbitrary sequence of numbers may not correspond to a feasible sequence of moments. We define $\mathcal{M}(k, \Omega)$ as the set of sequences of moments $(m_0, m_1, ..., m_k)$ feasible in Ω.[7] The problem of deciding whether or not a sequence of numbers is a feasible sequence of moments is called the classical moment problem [33]. Depending on the set Ω, we find three important instances of this problem:

(i) the Hamburguer moment problem, when $\Omega = \mathbb{R}$,

(ii) the Stieltjes moment problem, when $\Omega = \mathbb{R}_+$, and

(iii) the Hausdorff moment problem, where $\Omega = [0, 1]$.

For univariate distributions, necessary and sufficient conditions for feasibility of the classical moment problem can be given in terms of certain matrices being positive semidefinite. Let us define, for any $l \geq 0$, the following Hankel matrices of moments,

$$R_{2l} = \begin{bmatrix} m_0 & m_1 & \cdots & m_l \\ m_1 & m_2 & \cdots & m_{l+1} \\ \vdots & \vdots & \ddots & \vdots \\ m_l & m_{l+1} & \cdots & m_{2l} \end{bmatrix} \succeq 0, \text{ and } R_{2l+1} = \begin{bmatrix} m_1 & m_2 & \cdots & m_{l+1} \\ m_2 & m_3 & \cdots & m_{l+2} \\ \vdots & \vdots & \ddots & \vdots \\ m_{l+1} & m_{l+2} & \cdots & m_{2l+1} \end{bmatrix} \succeq 0. \tag{17}$$

[7] In what follows, we assume that $m_0 = 1$.

Hence, we have the following feasibility results for the Hamburguer and Stieltjes moment problems [33]:[8]

Theorem V.1 *Necessary and sufficient conditions for a sequence of moments* $\mathbf{m}_k = (m_0, m_1, ..., m_k)$ *to be feasible are:*

(a) $R_{2\left\lfloor \frac{k}{2} \right\rfloor} \succeq 0$*, for the Hamburguer moment problem* ($\Omega = \mathbb{R}$)*,*

(b) $R_{2\left\lfloor \frac{k}{2} \right\rfloor} \succeq 0$ *and* $R_{2\left\lfloor \frac{k+1}{2} \right\rfloor - 1} \succeq 0$*, for the Stieltjes moment problem* ($\Omega = \mathbb{R}_+$)*.*

Notice that, the spectral density $\mu_{\mathcal{G}}$ of a finite graph $\mathcal{G}$ is always supported by a bounded interval in $\mathbb{R}$; hence, its sequence of spectral moments $(1, m_1(A_{\mathcal{G}}), ..., m_5(A_{\mathcal{G}}))$ always satisfy the Hamburguer feasibility condition, $R_4 \succeq 0$. In the following subsection, we present the dual formulation proposed in [31] and [32] to compute the solution of (16) by solving a single semidefinite program.

B. SDP Dual Formulation

Using duality theory, one can associate a dual variable y_i to each equality constraint of the primal (16) to obtain [32]:

$$\begin{aligned} Z_D = \ & \min_{y_i} \ \sum_{i=0}^{k} y_i m_i \\ & \text{s.t. } \ p(x) = \sum_{i=0}^{k} y_i x^i \geq 1, \ \text{for } x \in T, \\ & \qquad p(x) = \sum_{i=0}^{k} y_i x^i \geq 0, \ \text{for } x \in \Omega. \end{aligned} \tag{18}$$

Notice that the dual constrains are univariate polynomials in x. Since a univariate polynomial is nonnegative if and only if it can be written as sum of squares of polynomials, the dual problem is a sum-of-squares program (SOSP) and can be formulated as a semidefinite program (using, for example, SOSTOOLS [34]). Karlin and Isii proved the following result concerning strong duality (see [35], p. 472):

Theorem V.2 *If the moment sequence* $(m_0, m_1, ..., m_k)$ *is in the interior of* $\mathcal{M}(k, \Omega)$*, then* $Z_P = Z_D$*.*

[8] Conditions for the Hausdorff moment problem to be feasible can also be found in [33], but they are not relevant in this work.

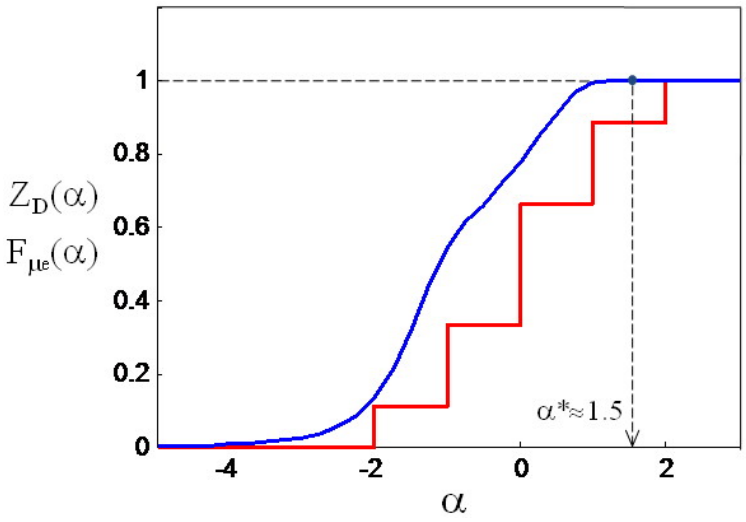

FIG. 8: Numerical solution of the SOSP described in Example V.1. The stairs-like function corresponds to the cumulative density function of $\mu_e(x)$. The (numerical) function $Z_D(\alpha)$ grows monotonically with α until it reaches a value $Z_D(\alpha^*) = 1$ at $\alpha^* \approx 1.5$.

From Theorem V.1, we have that a feasible sequence of moments, $\mathbf{m}_k \in \mathcal{M}(k, \Omega)$, is in the interior of $\mathcal{M}(k, \Omega)$ if $R_{2\lfloor \frac{k}{2} \rfloor} \succ 0$. Hence, for sequences of spectral moments $\mathbf{m}_k \notin \text{int}(\mathcal{M}(k, \mathbb{R}))$, we have that $\det R_{2\lfloor k/2 \rfloor} = 0$. Since the condition $\det R_{2\lfloor k/2 \rfloor} = 0$ is only satisfied for very degenerate networks, strong duality is almost always satisfied for real networks.

Example V.1 *In this example, we solve the SOSP in (18) for a given sequence of five moments,* $(m_k)_{k=1}^5 = (0, 4/3, 0, 4, 0)$, *when* $\Omega = [-10, 10]$ *and* $T = [-10, \alpha]$, *with* $\alpha = [-5 : 0.25 : 3]$. *The given set of moments corresponds to the atomic distribution* $\mu_e(x) = \sum_{i=1}^5 w_i \delta(x - x_i)$, *with support on the set of points* $(x_i)_{i=1}^5 = (-2, -1, 0, 1, 2)$ *and associated weights* $(w_i)_{i=1}^5 = (1/9, 2/9, 3/9, 2/9, 1/9)$. *We denote by* $Z_D(\alpha)$ *the solution to the SOSP as a function of the parameter* α, *and* $F_{\mu e}(\alpha) = \int_{-\infty}^{\alpha} \mu_e(x)\, dx$, *i.e., the cumulative distribution of* $\mu_e(x)$. *In Fig. 8, we plot both* $Z_D(\alpha)$ *and* $F_{\mu e}(\alpha)$ *for* $\alpha \in [5, 3]$.

Notice that $Z_D(\alpha)$ uniformly upper-bounds $F_{\mu e}(\alpha)$, since $Z_D(\alpha) = \max_{\mu} \int_{-\infty}^{\alpha} d\mu(x) \geq \int_{-\infty}^{\alpha} \mu_e(x)\, dx = F_{\mu e}(\alpha)$. In general, $Z_D(\alpha)$ uniformly upper bounds the cumulative distribution function of any random variable with the given sequence of moments. In Fig. 8, we also observe how the (numerically computed) function $Z_D(\alpha)$ monotonically increases

until it reaches the value 1 at $\alpha = \alpha^*$. In the next subsection, we shall prove that α^* is the largest lower bound on the spectral radius of $A_{\mathcal{G}}$ given a sequence of spectral moments. It is important to remark that the technique described above presents serious numerical problems from both ill-conditioning and numerical inefficiencies ([36], Lec. 12). Hence, the computation of the critical value α^* can run into numerical difficulties if we use the numerical sweep approach used in Example V.1. In Subsection V D, we shall introduce an alternative approach that allows us to accurately find the value of α^* directly from the roots of a univariate polynomial equation.

C. Bounds on the Spectral Radius from Spectral Moments

In this subsection, we show that any value of α for which $Z_D(\alpha) < 1$ is a lower bound on the spectral radius of $A_{\mathcal{G}}$. First, it is convenient to introduce several definitions. For a feasible moment sequence $\mathbf{m}_k = (1, m_1, m_2, ..., m_k) \in \mathcal{M}(k, \Omega)$, we define $\mathbb{M}(\mathbf{m}_k)$ as the set of positive Borel measures supported by $\mathbb{R}$ with moments matching those in $\mathbf{m}_k$. We also define $\mathbb{S}(\mathbf{m}_k) \subseteq \mathbb{M}(\mathbf{m}_k)$ as the set of spectral densities $\mu_{\mathcal{G}} = \frac{1}{n}\sum_{i=1}^{n} \delta(x - \lambda_i)$ satisfying $\frac{1}{n}\sum_{i=1}^{n} \lambda_i^j = m_j$, for $j = 1, ..., k$. We denote by $F_{\mu_{\mathcal{G}}}$ the cumulative distribution function of the spectral density $\mu_{\mathcal{G}} \in \mathbb{S}(\mathbf{m}_k)$, i.e.,

$$F_{\mu_{\mathcal{G}}}(\alpha) = \int_{-\infty}^{\alpha} \mu_{\mathcal{G}}(x)\, dx = \frac{1}{n}\left|\{\lambda_i : \lambda_i \leq \alpha\}\right|.$$

Furthermore, for a given sequence of moments $\mathbf{m}_k$, we define the function $Z_D(\alpha)$ as

$$Z_D(\alpha) \triangleq \max_{\mu \in \mathbb{M}(\mathbf{m}_k)} \int_{-\infty}^{\alpha} 1\, d\mu.$$

This function can be numerically computed following the technique used in Example V.1. It is easy to prove that $Z_D(\alpha)$ presents the following properties:

(i) $Z_D(\alpha)$ is monotone non-decreasing and right-continuous,

(ii) $\lim_{\alpha \to -\infty} Z_D(\alpha) = 0$ and $\lim_{\alpha \to \infty} Z_D(\alpha) = 1$, and

(iii) $Z_D(\alpha)$ uniformly upper bound $F_{\mu_{\mathcal{G}}}(\alpha)$ for any $\mu_{\mathcal{G}} \in \mathbb{S}(\mathbf{m}_k)$.

Properties (i) and (ii) indicate that $Z_D(\alpha)$ is a cumulative distribution function and, since $\mathbb{S}(\mathbf{m}_k) \subseteq \mathbb{M}(\mathbf{m}_k)$, Property (iii) comes from

$$Z_D(\alpha) = \max_{\mu \in \mathbb{M}(\mathbf{m}_k)} \int_{-\infty}^{\alpha} 1 \, d\mu \geq \max_{\mu_{\mathcal{G}} \in \mathbb{S}(\mathbf{m}_k)} \int_{-\infty}^{\alpha} 1 \, d\mu_{\mathcal{G}} \geq F_{\mu_{\mathcal{G}}}(\alpha).$$

For a given graph $\mathcal{G}$, we introduced in (5) an upper bound on the spectral radius of $A_{\mathcal{G}}$, denoted by $u_3 \geq \rho(A_{\mathcal{G}})$. This implies that, for the spectral density $\mu_{\mathcal{G}} \in \mathbb{S}(\mathbf{m}_k)$ associated to $\mathcal{G}$, we have that

$$F_{\mu_{\mathcal{G}}}(u_3) = \frac{1}{n} |\{\lambda_i : \lambda_i \leq u_3\}| = 1,$$

since $\lambda_i \leq u_3$ for all i. Therefore, since $0 \leq F_{\mu_{\mathcal{G}}}(\alpha) \leq Z_D(\alpha) \leq 1$, for any $\mu_{\mathcal{G}} \in \mathbb{S}(\mathbf{m}_k)$, there exists a value $c \leq u_3$ such that $Z_D(c) = 1$. In what follows, we propose an efficient technique to find the minimum value at which the c.d.f. $Z_D(\alpha)$ reaches the value one.

For a given sequence of five moments, $\mathbf{m}_5 = (1, m_1, ..., m_5)$, we define α^* as

$$\begin{aligned} \alpha^* &\triangleq \min\{\alpha \in \mathbb{R} : Z_D(\alpha) = 1\} \\ &= \sup\{\alpha \in \mathbb{R} : Z_D(\alpha) < 1\}. \end{aligned} \tag{19}$$

Hence, since $Z_D(\alpha^*) = 1$, α^* is the solution to the following optimization problem

$$\begin{aligned} \alpha^* = \ & \min_{\alpha,\mu} \ \alpha \\ & \text{s.t.} \ \int_{-\infty}^{\alpha} d\mu = 1, \\ & \quad \int_{-\infty}^{\alpha} x^j d\mu = m_j, \ j = 1, ..., 5. \end{aligned} \tag{20}$$

This problem is equivalent to the following feasibility moment problem:

Problem 2 *Find the minimum value of α for which there exists a positive Borel measure μ supported by $\Omega = (-\infty, \alpha]$ matching the sequence of moments $\mathbf{m}_5$.*

In the following subsection, we shall show how to efficiently compute the optimal value α^* by adapting Stieltjes' feasibility results. Our interest in α^* is due to the fact that it is the largest lower bound on $\rho(A_{\mathcal{G}})$ that can be found from a given sequence of spectral moments, as stated in the following Lemma:

Lemma V.3 *Consider a simple graph $\mathcal{G}$ with a sequence of spectral moments $\mathbf{m}_k = (1, m_1(A_{\mathcal{G}}), ..., m_k(A_{\mathcal{G}})) \in \text{int}(\mathcal{M}(k, \mathbb{R}))$. Hence, the spectral radius of $A_{\mathcal{G}}$ satisfies $\rho(A_{\mathcal{G}}) \geq \alpha^*$.*

Proof. For any $\alpha < \alpha^*$, we have that

$$1 > Z_D(\alpha) = \max_{\mu \in \mathbb{M}(\mathbf{m}_k)} \int_{-\infty}^{\alpha} 1 \, d\mu \geq \max_{\mu_{\mathcal{G}} \in \mathbb{S}(\mathbf{m}_k)} \int_{-\infty}^{\alpha} 1 \, d\mu_{\mathcal{G}}.$$

Hence, we have that

$$0 < 1 - \max_{\mu_{\mathcal{G}} \in \mathbb{S}(\mathbf{m}_k)} \int_{-\infty}^{\alpha} 1 \, d\mu_{\mathcal{G}} = \min_{\mu_{\mathcal{G}} \in \mathbb{S}(\mathbf{m}_k)} \int_{x>\alpha} 1 \, d\mu_{\mathcal{G}}(x).$$

Then, for any $\mu_{\mathcal{G}} \in \mathbb{S}(\mathbf{m}_k)$ with discrete support $\{\lambda_i\}_{i=1}^n$, we have

$$\int_{x>\alpha} 1 \, d\mu_{\mathcal{G}}(x) = \frac{1}{n} |\{\lambda_i : \lambda_i > \alpha\}| > 0.$$

Therefore, since $|\{\lambda_i : \lambda_i > \alpha\}| > 0$, we have that $\max \lambda_i = \rho(A_{\mathcal{G}}) > \alpha$, for any $\alpha < \alpha^*$. Furthermore, from (19), we have that $\rho(A_{\mathcal{G}}) \geq \alpha^*$. ■

Following the approach in Example V.1, the value of α^* can be numerically computed for $\mathbf{m}_k \in \text{int}(\mathcal{M}(k, \mathbb{R}))$ via a numerical sweep. As we remarked before, this SDP-based approach is ill-conditioned and presents numerical inefficiencies that make the accurate computation of α^* complicated and costly. In the following subsection, we present an efficient technique to compute the value α^* directly from the roots of a univariate polynomial equation.

D. Largest Lower Bound from Moments

Consider a graph $\mathcal{G}$ with sequence of spectral moments $\mathbf{m}_5 = (1, m_1(A_{\mathcal{G}}), ..., m_5(A_{\mathcal{G}}))$. In this section we show how to efficiently compute the value of α^* by studying the optimization problem in (20). Note that this problem is equivalent to finding the minimum value of α for which the sequence of moments $\mathbf{m}_5$ is feasible in the support $\Omega = (-\infty, \alpha]$. In what follows, we build on Stieltjes' feasibility conditions to solve this problem. The following lemma is useful in our derivations:

Lemma V.4 *Necessary and sufficient conditions for existence of a positive Borel measure μ supported by $\Omega = (-\infty, \alpha]$ matching the sequence of moments $\mathbf{m}_5 = (1, m_1, ..., m_5)$ are*

$$\tilde{R}_4 = \begin{bmatrix} \tilde{m}_0 & \tilde{m}_1 & \tilde{m}_2 \\ \tilde{m}_1 & \tilde{m}_2 & \tilde{m}_3 \\ \tilde{m}_2 & \tilde{m}_3 & \tilde{m}_4 \end{bmatrix} \succeq 0, \text{ and } \tilde{R}_5 = \begin{bmatrix} \tilde{m}_1 & \tilde{m}_2 & \tilde{m}_3 \\ \tilde{m}_2 & \tilde{m}_3 & \tilde{m}_4 \\ \tilde{m}_3 & \tilde{m}_4 & \tilde{m}_5 \end{bmatrix} \succeq 0, \tag{21}$$

where $\tilde{m}_k = \sum_{r=0}^{k} \binom{k}{r} (-1)^r \alpha^{k-r} m_r$.

Proof. Notice that the existence of a positive Borel measure $\mu \in \mathcal{M}(5, (-\infty, \alpha])$ is equivalent to the existence of a probability density function $\tilde{\mu}(x) = \mu(\alpha - x)$ supported by $\Omega = [0, \infty)$. The moments of $\tilde{\mu}$ are:

$$\begin{aligned}\tilde{m}_k &:= \int_{-\infty}^{\infty} x^k d\tilde{\mu}(x) = \int_{-\infty}^{\infty} (\alpha - x)^k d\mu(x) \\ &= \int_{-\infty}^{\infty} \sum_{r=0}^{k} \binom{k}{r} (-1)^r \alpha^{k-r} x^r d\mu(x) \\ &= \sum_{r=0}^{k} \binom{k}{r} (-1)^r \alpha^{k-r} m_r.\end{aligned}$$

Thus, applying Stieltjes feasibility conditions in Theorem V.1 to the above sequence of moments, we obtain the statement of Lemma V.4. ■

In order to solve (20), we need to find necessary and sufficient conditions under which $\tilde{R}_4$ and $\tilde{R}_5$ are positive semidefinite. The following property of $\tilde{R}_5$ is useful in our derivations:

Claim V.2 *For* $\mathbf{m}_5 = (1, m_1, ..., m_5) \in \text{int}(\mathcal{M}(5, \mathbb{R}))$ *and* $\tilde{R}_5(\alpha)$ *defined in (21), we have that*

$$\det \tilde{R}_5(\alpha) = \det R_4 \ \left(\alpha^3 + \alpha^2 c_2 + \alpha c_1 + c_0\right), \tag{22}$$

where c_0, c_1, *and* c_2 *are the solution of the following linear system of equations*

$$\begin{bmatrix} 1 & m_1 & m_2 \\ m_1 & m_2 & m_3 \\ m_2 & m_3 & m_4 \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ c_2 \end{bmatrix} = - \begin{bmatrix} m_3 \\ m_4 \\ m_5 \end{bmatrix}. \tag{23}$$

Proof. In Appendix II. ■

Remark V.1 *Since* $\mathbf{m}_5 \in \text{int}(\mathcal{M}(5, \mathbb{R}))$, *we have from Theorem V.1 that* $R_4 \succ 0$, *which implies that* $\det R_4 > 0$. *Notice that* R_4 *is exactly the matrix in the left-hand side of equation (23); thus, there exists a unique solution for* c_0, c_1, *and* c_2. *Furthermore, the roots of* $\alpha^3 + \alpha^2 c_2 + \alpha c_1 + c_0$ *are all real and simple ([36], Lec. 12).*

The following theorem state a necessary and sufficient condition for the moment problem in (20) to be feasible:

Theorem V.5 *Consider a sequence of moments* $\mathbf{m}_5 = (1, m_1, ..., m_5) \in \text{int}\,(\mathcal{M}\,(5, \mathbb{R}))$*, and let* $\{x_1, x_2, x_3\}$ *be the roots of the univariate polynomial*

$$x^3 + c_2 x^2 + c_1 x + c_0 = 0, \tag{24}$$

with coefficients $\{c_0, c_1, c_2\}$ *defined by the solution of the linear system of equations in (23). Then, there exists a positive Borel measure* μ *supported by* $\Omega = (-\infty, \alpha]$ *matching the sequence of moments* $\mathbf{m}_5$ *if and only if* $\alpha \geq \max\{x_1, x_2, x_3\} := \alpha^*$.

Proof. From Lemma V.4, we need to find the set of values of α such that the following conditions are satisfied: (a) $\tilde{R}_4(\alpha) \succeq 0$ and (b) $\tilde{R}_5(\alpha) \succeq 0$ (where the entries of $\tilde{R}_4(\alpha)$ and $\tilde{R}_5(\alpha)$ are polynomials in α).

Condition a: Since $\mathbf{m}_5 \in \text{int}\,(\mathcal{M}\,(5, \mathbb{R}))$, there exists a positive measure μ supported by $\mathbb{R}$ matching the moments in $\mathbf{m}_5$. As shown in Lemma V.4, the moments of the measure $\tilde{\mu}(x) = \mu(\alpha - x)$ are $\tilde{m}_k = \sum_{r=0}^{k} \binom{k}{r} (-1)^r \alpha^{k-r} m_r$, where m_r is the r-th moment of μ. Clearly, $\tilde{\mu}$ is supported by $\mathbb{R}$; hence, the moment sequence $(1, \tilde{m}_1, ..., \tilde{m}_5)$ satisfy the Hamburguer feasibility conditions, i.e., $\tilde{R}_4(\alpha) \succeq 0$, for all $\alpha \in \mathbb{R}$.

Condition b: For a given sequence of moments, we have to find the set of values of α for which $\tilde{R}_5(\alpha) \succeq 0$. (Notice that, since $\tilde{R}_4(\alpha) \succeq 0$ for all values of α, the condition $\tilde{R}_5(\alpha) \succeq 0$ is necessary and sufficient for the existence of a positive Borel measure μ supported by $\Omega = (-\infty, \alpha]$ matching the sequence of moments $\mathbf{m}_5$.) The Hankel matrix $\tilde{R}_5(\alpha)$ satisfies the following two properties:

(**P1**) If $\tilde{R}_5(\alpha) \succeq 0$, then $\tilde{R}_5(\alpha') \succeq 0$ for all $\alpha' \geq \alpha$. This property is derived from the fact that, if $\mathbf{m}_5 \in \mathcal{M}\,(5, (-\infty, \alpha])$, then $\mathbf{m}_5 \in \mathcal{M}\,(5, (-\infty, \alpha'])$ for all $\alpha' \geq \alpha$.

(**P2**) If $\tilde{R}_5(\alpha) \not\succeq 0$, then $\tilde{R}_5(\alpha'') \not\succeq 0$ for all $\alpha'' \leq \alpha$. Similarly as above, if $\mathbf{m}_5 \notin \mathcal{M}\,(5, (-\infty, \alpha])$, then $\mathbf{m}_5 \notin \mathcal{M}\,(5, (-\infty, \alpha''])$ for all $\alpha'' \leq \alpha$.

Hence, from (**P1**) and (**P2**), we have that the set of values of α for which $\tilde{R}_5(\alpha) \succeq 0$ is of the sort $[\alpha^*, \infty)$, where α^* is the minimum value of α for which $\tilde{R}_5(\alpha) \succeq 0$. At $\alpha = \alpha^*$, we have that $\det \tilde{R}_5(\alpha^*) = 0$. From Claim V.2, we have that $\det \tilde{R}_5(\alpha) = \det R_4(\alpha) \ \ (\alpha^3 + \alpha^2 c_2 + \alpha c_1 + c_0) = 0$. Since $\det R_4(\alpha) \geq 0$, for all α, and the roots of $\alpha^3 + \alpha^2 c_2 + \alpha c_1 + c_0$ are all real and simple, we have that the minimum value of α for which $\tilde{R}_5(\alpha) \succeq 0$ is the maximum root of $\alpha^3 + \alpha^2 c_2 + \alpha c_1 + c_0$. ■

We now state our main result, with which we can efficiently compute the value the largest lower bound on the spectral radius of the adjacency matrix as the solution of a single polynomial equation:

Theorem V.6 *Consider a simple graph $\mathcal{G}$ with a sequence of spectral moments $\mathbf{m}_5 = (1, m_1, ..., m_5) \in \text{int}\left(\mathcal{M}\left(5, \mathbb{R}\right)\right)$. Then,*

$$\rho\left(A_{\mathcal{G}}\right) \geq \max\left\{x_1, x_2, x_3\right\}, \tag{25}$$

where $\{x_1, x_2, x_3\}$ are the roots of (24).

Proof. This result comes straightforward from Lemma V.3 and Theorem V.5. ■

Note that the roots x_1, x_2, x_3 in (25) can be written explicitly in terms of the coefficient c_0, c_1, and c_2, using Cardano's formula for cubic equations [37]. Also, these coefficients can be written explicitly in terms of the spectral moments $m_1, ..., m_5$ using Cramer's rule in (23). Furthermore, we derived explicit expressions for the spectral moments in terms of structural properties in Section IV. Therefore, we have explicit (but rather complicated) expressions for the optimal bound on the spectral radius in terms of structural properties. Since the spectral radius has a direct relationship with the spreading rate of a virus in the network, our expressions can be used to study the impact of the structural properties involved in expressions for the spectral moments in the dynamics of the viral spreading.

Example V.2 *Let us consider the sequence of moments in Example V.1. In that example, we estimated the value $\alpha^* \approx 1.5$ (see Fig. 8) using a costly and inefficient numerical sweep. Using Theorem V.5, we can efficiently and accurately compute α^* solving a single polynomial equation. In particular, from the sequence of moments $(m_k)_{k=1}^5 = (0, 4/3, 0, 4, 0)$, we have that the solution of the linear system of equations in (23) is $\{c_0 = 0, c_1 = -3, c_2 = 0\}$. Hence, the polynomial in (24) is equal to $x^3 - 3x$, and its roots are $\{x_1 = -\sqrt{3},\ x_2 = 0,\ x_3 = \sqrt{3}\}$. Therefore, we have that $\alpha^* = \sqrt{3} \approx 1.73$.*

Using the methodology introduced above, we are able to find optimal lower bounds on the spectral radius of $A_{\mathcal{G}}$ from a sequence of spectral moments of $\mathcal{G}$. In Section IV, we showed how to compute a sequence of five spectral moments from the aggregation of structural measurements. In the following section, we use our results to relate the global behavior of spreading processes in networks with local structural properties.

VI. NUMERICAL SIMULATIONS

In this section, we study viral spreading processes in online social networks. In particular, we analyze network data from the New Orleans regional network of Facebook [38]. The data was collected via a breadth-first-search (BFS) crawl of the network starting from a single user, and it comprises all friendship links among those users who made their profiles visible to the network. In total, the social graph spans $63,731$ users connected by $817,090$ undirected links. In order to corroborate our results in multiple social graphs, we extract multiple medium-size subgraphs from the available graph by using BFS around randomly chosen nodes. Each BFS scans the nodes around each chosen node, stopping when it explores all nodes 2 hops away. We then consider the subgraph resulting from the BFS as a different instance of a social graph, although this procedure introduces biases in the network properties and quantitative results must be considered carefully [39].[9]

In our first simulation, we validate the usage of the spectral radius as a measure of the spreading abilities of a social network. We consider 25 different social subgraphs, with sizes between 1000 and 3000 nodes, and study the behavior of viral spreading processes for values of δ/β equal to ρ, 0.85ρ, and 0.70ρ. For each value of δ/β, we simulate 25 different realizations of the viral spreading, with a 2% probability of initial infection, and compute the final fraction of individuals in the network that have been infected during the time of the simulation (250 steps). In Fig. 9, we plot the average fraction of infected individuals versus $\rho\left(A_{\mathcal{G}}\right)$ for the three different values of δ/β. We observe that for $\delta/\beta = \rho\left(A_{\mathcal{G}}\right)$ the portion of infected population lie under 20% for most cases (circles in Fig. 9). For $\delta/\beta = 0.70\rho\left(A_{\mathcal{G}}\right)$, the portion of infected mostly lies above 80% (squares in Fig. 9). For $\delta/\beta = 0.85\rho\left(A_{\mathcal{G}}\right)$, all the points lie in the transition band 20%-80%.

In our second numerical experiment, we consider a set of 100 social subgraphs, obtained via BFS exploration of second-order neighborhoods around randomly chosen nodes. We discard those subgraphs that are under 100 nodes and obtained a set of subgraphs from 100 to over $7,000$ nodes. For each social subgraph, we computed its first five spectral moments using its distribution of local measurements (via Corollary IV.2, and Lemmas IV.3 and IV.5). As a reminder, these measurements are the distributions of degrees, triangles, quadrangles

[9] Datasets and MATLAB code used in our simulations are available, in anonymized form, in [29].

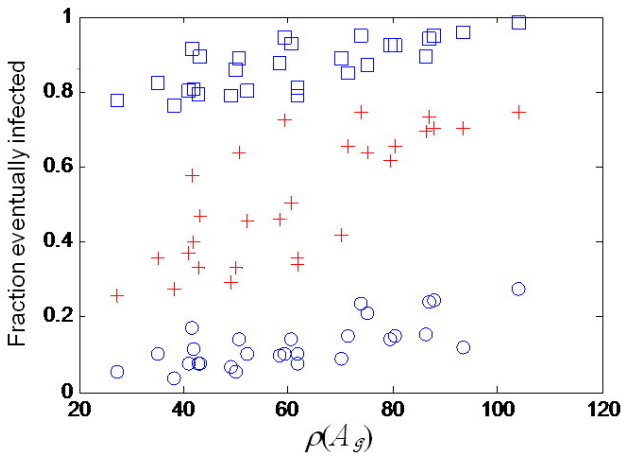

FIG. 9: Each point represent the average fraction of eventually infected nodes in 25 different social subgraphs versus the spectral radius, when δ/β equals ρ (circles), 0.85ρ (crosses), and 0.70ρ (squares).

and pentagons in the network, as well as the triangle-degree correlation $\mathcal{C}_{dt}$. Based on these moments, we also compute the lower bound on the spectral radius, α^*, by applying Theorem V.6. Although computing α^* is most useful when we cannot compute the spectral radius of the network, our numerical setup allows us to compute the exact value of the spectral radius, which we have included for comparison purposes. Some of the numerical results in this experiment are included in Table 1.

n_i	m_1	m_2	...	m_5	α^*	ρ	α^*/ρ
n_1=5,866	0	45.6	...	3.48e6	69.5	102.4	0.68
n_2=3,940	0	53.5	...	7.64e6	85.9	116.1	0.74
⋮	⋮	⋮	⋱	⋮	⋮	⋮	⋮
n_{100}=694	0	23.7	...	5.64e5	42.9	50.4	0.85

TABLE 1. Each row in the above table corresponds to a different social subgraph. In each row, we include the sizes (n_i), moments (m_j), the lower bound on the spectral radius (α^*), the exact value of $\rho(A_{\mathcal{G}})$, and the ratio $\alpha^*/\rho(A_{\mathcal{G}})$.

In Fig. 10, we include a scatter plot where each point is associated with each one of the 100 social subgraphs considered in our experiments. The coordinates of each point corresponds to the spectral radius, $\rho(A_{\mathcal{G}})$, and its lower bound, α^*, respectively. We observe how *the*

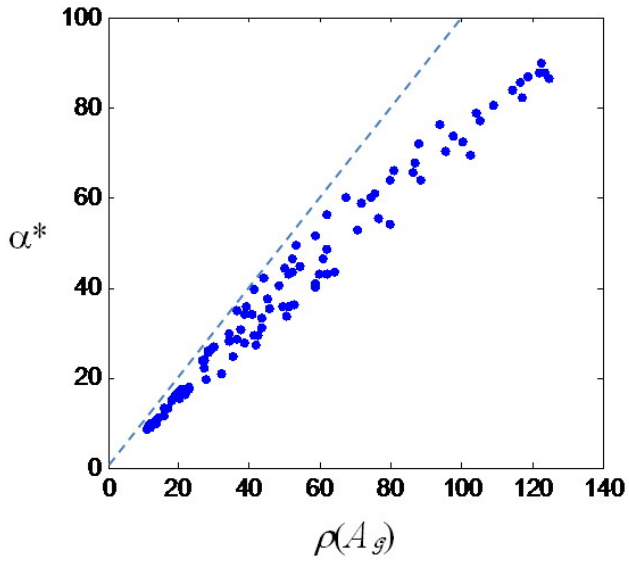

FIG. 10: Scatter plot of the spectral radius, $\rho(A_{\mathcal{G}})$, versus its lower bound, α^*, where each point is associated with each one of the 100 social subgraphs considered in our experiments.

spectral radii of these social subgraphs are remarkably close to the theoretical lower bound, specially for small values of $\rho(A_{\mathcal{G}})$. This empirical observation validates the usage of α^* as an estimator of the spectral radius of a network based on its structural properties.

In Fig. 11, we plot the ratio $\rho(A_{\mathcal{G}})/\alpha^*$ with respect to the network size (in log scale). We observe how this ratio is very close to one for social subgraphs ranging from 10^2 to 10^3 nodes, and it decreases as we increase the subgraph size towards 10^4 nodes. This indicates that, as we increase the network size, richer structural measurements are needed to accurately determine the network spreading abilities.

VII. CONCLUSIONS

A fundamental question in the field of network science is to understand the relationship between a network structure and its dynamical performance. The common approach to study this relationship is by means of synthetic network models. This approach present a major flaw: Synthetic network models implicitly induce structural properties that are not directly controlled and can be relevant to the network dynamical performance. Therefore, it is difficult to isolate the role of a particular network property in the dynamical performance using synthetic networks.

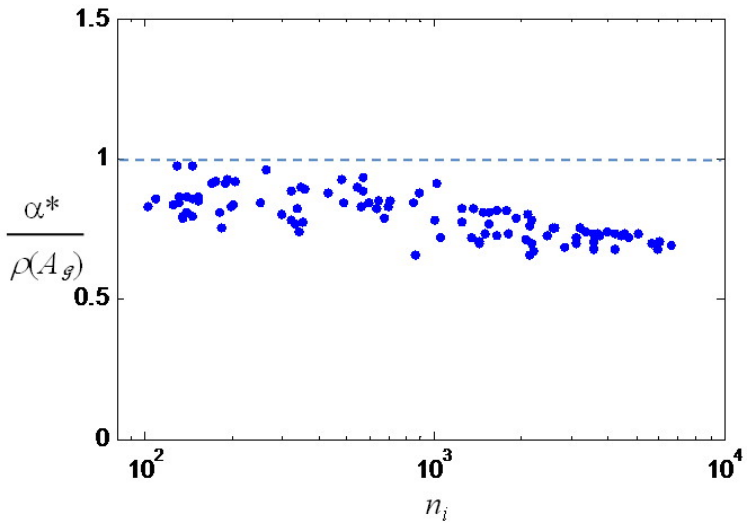

FIG. 11: Plot of $\alpha^*/\rho\,(A_{\mathcal{G}})$ with respect to the network size (in log scale) for the 100 social subgraphs considered in our experiments.

In this paper, we have proposed an alternative approach to study the relationship between structure and dynamics that does not use synthetic models. We have illustrated our approach by studying the dynamics of the susceptible-infected-susceptible model of viral propagation. In particular, we have showed how structural properties constrain the rate of spreading of a virus in the network. In our analysis, we have exploited the close relationship between the spectral radius of the adjacency matrix of the network and the expected rate of spreading of a random viral infection. Our work builds on algebraic graph theory to find expressions for the spectral moments of the adjacency matrix in terms of structural network properties. We have then used convex optimization techniques to find optimal bounds on the spectral radius from a truncated sequence of spectral moments. Finally, we have used these bounds to study the relationship between network structural properties and the dynamical performance of spreading processes in real instances of online social networks.

Our approach explicitly unveils what structural properties are most relevant to the network dynamical performance. In particular, we show that not only the distribution of edges and triangles, but also their correlation and the presence of larger cycles, are essential in the performance of spreading processes. In our analysis, we have found expressions that quantify the impact of these structural features on the spectral moments. These expressions allow us to make an informed decision on what structural properties are most relevant for a network's ability to spread information virally. Therefore, to study the effect of a particular

structural property in the network dynamical performance, we must use models in which we can jointly and explicitly control all the relevant structural properties.

APPENDIX A: PROOF OF LEMMA IV.5

Lemma IV.5 Let $\mathcal{G}$ be a simple graph. Denote by p_i, t_i, and d_i the number of pentagons, triangles, and edges touching node i in $\mathcal{G}$, respectively. Then,

$$m_5\left(A_{\mathcal{G}}\right) = \frac{1}{n}\left[\sum_{i=1}^{n} 2p_i + 10t_i d_i - 10t_i\right].$$

Proof. As in Lemma (IV.3), we count the number of closed walks of length 5 of different types. We provided a classification of the walk types in Fig. 3, where we also included expressions for the number of closed walks of each type in terms of local metrics). We now provide the details on how to compute those expressions for each walk type:

Type (a) The number of closed walks of this type starting at i is equal to twice the number of pentagons touching i, hence, the total number is given by $\sum_{i=1}^{n} 2p_i$.

Type (b) In order to count walks of this type, it is convenient to define t_{pqr} as the indicator function that takes value 1 if there exists a triangle among vertices p, q, and r; 0 otherwise. Note that this indicator satisfies $\sum_{q=1}^{n}\sum_{r=1}^{n} t_{pqr} = 2t_p$. Hence, the number of closed walks in this type can be written as:

$$w_5^{(b)} = \sum_{i=1}^{n}\sum_{p=1}^{n}\sum_{q\neq i}\sum_{r\neq i} a_{ip}t_{pqr},$$

where a_{ip} indicates the existence of an edge from i to p, and t_{pqr} indicates the existence of a triangle for q and r different than i (we impose this inequalities so that the underlying graph does not become a triangle). We can then perform the following algebraic manipulations,

$$\begin{aligned} w_5^{(b)} &\overset{(i)}{=} \sum_{p=1}^{n}\sum_{q=1}^{n}\sum_{r=1}^{n} t_{pqr}\left(\sum_{i\neq q,r} a_{ip}\right) \\ &\overset{(ii)}{=} \sum_{p=1}^{n}\left(d_p - 2\right)\sum_{q=1}^{n}\sum_{r=1}^{n} t_{pqr} \\ &= 2\sum_{p=1}^{n} t_p\left(d_p - 2\right), \end{aligned}$$

where in equality (i) we have changed the order of the subindices, and impose the inequality constrains on subindex i (note that this change does not affect the set of permutation of subindices that satisfy the inequalities). In equality (ii), we take into account that $\sum_{i\neq q,r} a_{ip} = \sum_{i=1}^{n} a_{ip} - a_{qp} - a_{rp} = d_i - 2$, since p is connected to q and r in this walk type.

Type (c) We can use the indicator function t_{ijk} to write the total number of walks in this type as follows,

$$\begin{aligned} w_5^{(c)} &= 2\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} t_{ijk}\,(d_j - 2) \\ &= 4\sum_{j=1}^{n} (d_j - 2)\, t_j, \end{aligned}$$

where the last expression comes from reordering the summations and applying $\sum_{i=1}^{n}\sum_{k=1}^{n} t_{ijk} = 2t_j$.

Type (d) The number of walks starting at i in this type is equal to $4t_i\,(d_i - 2)$, where we have included a -2 in the parenthesis to take into account that two of the edges touching i are part of the triangle. The coefficient 4 accounts for the two possible direction we can walk the triangle and the two possible choices for the first step of the walk (towards the triangle or towards the single edge).

Type (e) and (f) These types of walks correspond to the set of closed walks of length 5 that visit all (and only) the edges of a triangle. Given a particular triangle touching i, we can easily count the number of walks of this type to be equal to 10, where 8 of them are of type (e) and 2 of type (f).

Hence, we obtain (10) by summing up all the above contributions (and simple algebraic manipulations). ■

APPENDIX B: PROOF OF CLAIM V.2

In this proof, we make use of Vandermonde's convolution formula:

$$\binom{m+n}{r} = \sum_{k=0}^{r} \binom{m}{k}\binom{n}{r-k},$$

for $m, n, r \in \mathbb{N}_0$.

Claim V.2 *For* $m_5 = (1, m_1, ..., m_5) \in \operatorname{int}(\mathcal{M}(5, \mathbb{R}))$ *and* $\tilde{R}_5(\alpha)$ *defined in (21), we have that*

$$\det \tilde{R}_5(\alpha) = \det R_4 \ \left(\alpha^3 + \alpha^2 c_2 + \alpha c_1 + c_0\right),$$

where c_0, c_1, *and* c_2 *are the solution of the following linear system of equations*

$$\begin{bmatrix} 1 & m_1 & m_2 \\ m_1 & m_2 & m_3 \\ m_2 & m_3 & m_4 \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ c_2 \end{bmatrix} = - \begin{bmatrix} m_3 \\ m_4 \\ m_5 \end{bmatrix}.$$

Proof. First, we define the matrix

$$C = \begin{bmatrix} 1 & \tilde{m}_1 & \tilde{m}_2 & \tilde{m}_3 \\ \tilde{m}_1 & \tilde{m}_2 & \tilde{m}_3 & \tilde{m}_4 \\ \tilde{m}_2 & \tilde{m}_3 & \tilde{m}_4 & \tilde{m}_5 \\ -1 & 0 & 0 & 0 \end{bmatrix} \tag{B1}$$

where $\tilde{m}_k = \sum_{r=0}^{k} \binom{k}{r} (-1)^r \alpha^{k-r} m_r$. Notice that $\det \tilde{R}_5 = \det C$. Second, we pre- and post-multiply C by 4-by-4 lower and upper triangular matrices, $L = [l_{ij}]$ and $U = [u_{ij}]$, defined entry-wise as:

$$l_{ij} = \begin{cases} \binom{i-1}{j-1} \alpha^{i-j} (-1)^{j-1}, & \text{for } 1 \le i \le 3,\ j \le i, \\ 1, & i = j = 4 \\ 0, & \text{otherwise} \end{cases}$$

$$u_{ij} = \begin{cases} \binom{j-1}{i-1} (-1)^{i-1} \alpha^{j-i}, & \text{for } 1 \le i \le 4,\ j \ge i, \\ 0, & \text{otherwise.} \end{cases}$$

Since both matrices have ones in their diagonals, we have that $\det C = \det(LCU)$.

We can use Vandermonde's convolution formula and exploit the structure of L and U to explicitly compute the entries of (LCU) as follows. For $0 \le i \le 2$, we can expand the matrix

multiplication as

$$\begin{aligned}(LCU)_{i+1,j+1} &= \sum_{r=1}^{i+1}\sum_{s=1}^{j+1} l_{i+1,r}C_{r,s}u_{s,j+1} \\ &= \sum_{r=0}^{i}\sum_{s=0}^{j} l_{i+1,r+1}C_{r+1,s+1}u_{s+1,j+1} \\ &= \sum_{r=0}^{i}\sum_{s=0}^{j} \binom{i}{r}\binom{j}{s}(-1)^{r+s}\,\alpha^{(i+j)-(r+s)}\tilde{m}_{r+s}\end{aligned}$$

where we use the fact that $C_{r+1,s+1} = \tilde{m}_{r+s}$ for $0 \leq r \leq 2$, in the last equality. Performing the change of variables $k = r + s$, we can rewrite the above summations in terms of r and k as

$$\begin{aligned}(LCU)_{i+1,j+1} &= \sum_{k=0}^{i+j}\sum_{r=0}^{k} \binom{i}{r}\binom{j}{k-r}(-1)^{k}\,\alpha^{(i+j)-k}\tilde{m}_{k} \\ &= \sum_{k=0}^{i+j}(-1)^{k}\,\alpha^{(i+j)-k}\tilde{m}_{k}\left[\sum_{r=0}^{k}\binom{i}{r}\binom{j}{k-r}\right] \\ &\overset{(a)}{=} \sum_{k=0}^{i+j}\binom{i+j}{k}(-1)^{k}\,\alpha^{(i+j)-k}\tilde{m}_{k} \overset{(b)}{=} m_{i+j},\end{aligned}$$

where we have used Vandermonde's convolution formula in equality (a), and the fact that $m_t = \sum_{s=0}^{t}\binom{t}{s}(-1)^{s}\,\alpha^{t-s}\tilde{m}_s$ in equality (b). Hence, the first three rows of (LCU) present a Hankel-like structure containing the moments m_k, $k = 0, ..., 5$. Also, from the structure of L and C, we have that the last row of (LCU) is equal to $(LCU)_{4,j+1} = -u_{1,j+1} = \alpha^{j}$. Hence, the structure of (LCU) is as follows

$$LCU = \begin{bmatrix} 1 & m_1 & m_2 & m_3 \\ m_1 & m_2 & m_3 & m_4 \\ m_2 & m_3 & m_4 & m_5 \\ 1 & \alpha & \alpha^2 & \alpha^3 \end{bmatrix}.$$

Performing a Laplace expansion along the last row of LCU, and permuting some of the

columns of the cofactors, we have

$$\det(LCU) = \det\begin{bmatrix} -m_3 & m_1 & m_2 \\ -m_4 & m_2 & m_3 \\ -m_5 & m_3 & m_4 \end{bmatrix} + \alpha \det\begin{bmatrix} 1 & -m_3 & m_2 \\ m_1 & -m_4 & m_3 \\ m_2 & -m_5 & m_4 \end{bmatrix}$$
$$+ \alpha^2 \det\begin{bmatrix} 1 & m_1 & -m_3 \\ m_1 & m_2 & -m_4 \\ m_2 & m_3 & -m_5 \end{bmatrix} + \alpha^3 \det\begin{bmatrix} 1 & m_1 & m_2 \\ m_1 & m_2 & m_3 \\ m_2 & m_3 & m_4 \end{bmatrix},$$

Based on Cramer's rule, we can rewrite the determinant of the above cofactors as

$$\det(LCU) = \alpha^3 \det R_4 + \alpha^2 c_2 \det R_4 + \alpha c_1 \det R_4 + c_0 \det R_4,$$

where c_0, c_1, and c_2 are the solution of the linear of equations

$$\begin{bmatrix} 1 & m_1 & m_2 \\ m_1 & m_2 & m_3 \\ m_2 & m_3 & m_4 \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ c_2 \end{bmatrix} = - \begin{bmatrix} m_3 \\ m_4 \\ m_5 \end{bmatrix},$$

as we wanted to prove. ■

---

[1] S.H. Strogatz, "Exploring Complex Networks," *Nature*, vol. 410, pp. 268-276, 2001.

[2] A.L. Barabási, and R. Albert, "Emergence of Scaling in Random Networks," *Science*, vol. 285, pp. 509-512, 1999.

[3] D.J. Watts and S. Strogatz, "Collective Dynamics of Small World Networks," *Nature,* vol 393, pp. 440-42, 1998.

[4] M. Molloy and B. Reed, "A Critical Point for Random Graphs with a Given Degree Sequence," *Random Structures and Algorithms*, vol. 6, pp. 161-180, 1995.

[5] M.E.J. Newman, S.H. Strogatz, and D.J. Watts, "Random Graphs with Arbitrary Degree Distributions and Their Applications," *Physical Review E*, vol. 64, 026118, 2001.

[6] M.E.J. Newman, "Random Graphs with Clustering," *Physical Review Letters*, vol. 103, 058701, 2009.

[7] M.E.J. Newman, "Assortative Mixing in Networks," *Physical Review Letters*, vol. 89, 208701, 2002.

[8] A. Clauset, C. Moore, and M.E.J. Newman, "Hierarchical Structure and the Prediction of Missing Links in Networks," *Nature*, vol. 453, pp. 98-101, 2008.

[9] S. Boccaletti S., V. Latora, Y. Moreno, M. Chavez, and D.-H. Hwang, "Complex Networks: Structure and Dynamics," *Physics Reports*, vol. 424, no. 4-5, pp. 175-308, 2006.

[10] D. Alderson, L. Li, W. Willinger, and J.C. Doyle, "Understanding Internet Topology: Principles, Models, and Validation," *IEEE/ACM Transactions on Networking*, vol. 13, pp. 1205-1218, 2005.

[11] R.M. Anderson and R.M. Ray, *Infectious Diseases of Humans: Dynamics and Control*, Oxford University Press, 1991.

[12] M. Draief, A. Ganesh, and L. Massoulié, "Thresholds for Virus Spread on Networks," *Annals of Applied Probability*, vol. 18, pp. 359-378, 2008.

[13] D. Chakrabarti, Y. Wang, C. Wang, J. Leskovec, and C. Faloutsos, "Epidemic Thresholds in Real Networks," *ACM Transactions on Information and System Security*, vol. 10, no. 4, 2008.

[14] L. Zager and G.C. Verghese, "Epidemic Thresholds for Infections in Uncertain Networks," *Complexity*, vol. 14, pp. 12-25, 2009.

[15] M. Vojnović, V. Gupta, T. Karagiannis, and C. Gkantsidis, "Sampling Strategies for Epidemic-Style Information Dissemination," *IEEE/ACM Transactions on Networking*, vol. 28, pp. 1013-1025, 2010.

[16] J. Leskovec, L. Adamic, and B. Huberman, "The Dynamics of Viral Marketing," *ACM Transactions on the Web*, vol. 1, no. 5, 2007.

[17] L.M. Pecora and T.L. Carroll, "Master Stability Functions for Synchronized Coupled Systems," *Physical Review Letters*, vol. 80, pp. 2109-2112, 1998.

[18] C. Asavathiratham, S. Roy, B. Lesieutre, and G. Verghese, "The Influence Model," *IEEE Control Systems Magazine*, vol. 21, pp. 52-64, 2001.

[19] V.M. Preciado and G.C. Verghese, "Synchronization in Generalized Erdös-Rényi Networks of Nonlinear Oscillators," *Proc. IEEE Conference on Decision and Control*, pp. 4628-4633, 2005.

[20] V.M. Preciado, *Spectral Analysis for Stochastic Models of Large-Scale Complex Dynamical Networks*, Ph.D. dissertation, Dept. Elect. Eng. Comput. Sci., MIT, Cambridge, MA, 2008.

[21] V.M. Preciado and A. Jadbabaie, "Spectral Analysis of Virus Spreading in Random Geometric Networks," *Proc. IEEE Conf. on Decision and Control*, pp. 4802-4807, 2009.

[22] P. Van Mieghem, J. Omic, and R. Kooij, "Virus Spread in Networks," *IEEE/ACM Transac-*

tions on Networking, vol. 17 , no. 1, pp. 1-14, 2009.

[23] S.P. Meyn and R.L. Tweedie, *Markov Chains and Stochastic Stability*, Springer-Verlag, 1993.

[24] K.C. Das and P. Kumar, "Some New Bounds on the Spectral Radius of Graphs," *Discrete Mathematics*, vol. 281, pp. 149-161, 2004.

[25] O. Favaron, M. Maheo, and J.-F. Sacle, "Some Eigenvalue Properties in Graphs (Conjectures of Graffiti–II)," *Discrete Mathematics*, vol. 111, pp. 197-220, 1993.

[26] J. Shu and Y. Wu, "Sharp Upper Bounds on the Spectral Radius of Graphs," *Linear Algebra and its Applications*, vol. 377, pp. 241-248, 2004.

[27] N. Biggs, *Algebraic Graph Theory*, Cambridge University Press, 2^{nd} Edition, 1993.

[28] E. Ravasz, A.L. Somera, D.A. Mongru, Z. Oltvai, and A.-L. Barabási, "Hierarchical Organization of Modularity in Metabolic Networks", *Science,* vol. 297, pp. 1551-1555, 2002.

[29] <http://code.google.com/p/optimal-viral-bounds/>

[30] V.M. Preciado and A. Jadbabaie, "From Local Measurements to Network Spectral Properties: Beyond Degree Distributions," *Proc. IEEE Conference on Decision and Control*, 2010.

[31] J.B. Lasserre, "Bounds on Measures Satisfying Moment Conditions," *Annals of Applied Probability*, vol. 12, pp. 1114-1137, 2002.

[32] I. Popescu and D. Bertsimas, "An SDP Approach to Optimal Moment Bounds for Convex Classes of Distributions," *Mathematics of Operation Research*, vol. 50, pp. 632-657, 2005.

[33] J.A. Shohat and J.D. Tamarkin, *The Problem of Moments*, American Mathematical Society, 1943.

[34] S. Prajna, A. Papachristodoulou, P. Seiler, and P.A. Parrilo, "SOSTOOLS: Sum of Squares Optimization Toolbox for MATLAB," 2004. Available from <http://www.cds.caltech.edu/sostools>.

[35] S. Karlin and W.J. Studden, *Tchebycheff Systems: with Applications in Analysis and Statistics*, John Wiley and Sons, 1966.

[36] P. Parrilo, *Algebraic Techniques and Semidefinite Optimization*, Massachusetts Institute of Technology: MIT OpenCourseWare, Spring 2006.

[37] M. Abramowitz and I.A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Dover, 1965.

[38] B. Viswanath, A. Mislove, M. Cha, and K.P. Gummadi, "On the Evolution of User Interaction in Facebook," *Proc. ACM SIGCOMM Workshop on Social Networks*, 2009.

[39] M. Stumpf, C. Wiuf, and R. May, "Subnets of Scale-Free Networks are not Scale-Free: Sampling Properties of Networks," *Proceedings of the National Academy of Sciences*, vol. 102, pp. 4221-4224, 2005.