Predicting Network Centralities from Node Attributes

Harold Soh
Singapore-MIT Alliance for Research
and Technology (SMART)
Singapore 138602
haroldsoh@smart.mit.edu

Abstract

Among the variety of complex network metrics proposed, node importance or *centrality* has potentially found the most widespread application—from the identification of gene-disease associations to finding relevant pages in web search. In this workshop paper, we present a method that learns mappings from node attributes to latent centralities. We first construct an eigenvector-based Bayesian centrality model, which casts the problem of computing network centrality as one of probabilistic (latent variable) inference. Then, we develop the sparse variational Bayesian centrality Gaussian process (VBC-GP) which simultaneously infers the centralities and learns the mapping. The VBC-GP possesses inherent benefits: it (i) allows a potentially large number of nodes to be represented by the sparse mapping and (ii) permits prediction of centralities on previously unseen nodes. Experiments show that the VBC-GP learns high-quality mappings and compares favorably to a two-step method, i.e., a full-GP trained on the node attributes and network centralities. Finally, we present a case-study using the VBC-GP to distribute a limited number of vaccines to decrease the severity of a viral outbreak.

1 Introduction

Many real-world systems—from brain structures to the world-wide-web—are elegantly represented and analyzed as graphs or networks, i.e., nodes connected by links or edges. This representation lends itself easily to analysis using a diverse set of tools and metrics developed over the years.

That said, a majority of analyses are performed under the assumption of noise-free graphs. This is in opposition to reality where "clean" graphs are the exception, not the norm. Links are typically estimated from noisy data (e.g., measurements in wet-lab experiments), which lead to inaccurate estimates of network measures including centrality [1, 2, 3]. In particular, the eigenvector centrality measure undergoes a smooth decay in accuracy (relative to the centrality of the true network) as link errors increase [3].

In a related research thread, recent complex networks models proposed by the machine-learning community have involved node attributes (features) as determinants of edge formation. For example, the Multiplicative Attribute Graph (MAG) model [5] connects nodes probabilistically based on binary attributes and affinity matrices. In [6], Palla, Knowles and Ghahramani propose a hierarchical Bayesian model using latent node feature vectors to capture the probability of links. These models (and others) have found success in link prediction tasks and reproducing the connectivity patterns of synthetic and real-world networks.

In this work, we address the two issues above and ask the question: can network centralities be predicted from observed features and noisy network weights? We adopt a probabilistic approach by treating centrality estimation as probabilistic latent variable inference. We first construct a novel Bayesian centrality model for networks with noisy weighted links and then extend it via a sparse

Gaussian process (GP) to yield the variational Bayesian centrality Gaussian Process (VBC-GP). Unlike previous works into uncovering network centrality, the VBC-GP (i) directly addresses the problem of uncertainty when inferring node importance, (ii) permits principled assimilation of repeated weight observations and (iii) learns a mapping between observed node attributes to the latent centralities. This allows us to represent a potentially large number of nodes using a small number of inducing variables—latent function values at specific input locations that introduce conditional independence between the other latent variables—and to directly predict node centralities from node attributes.

Experiments on three real-world networks (with corrupted link weights) demonstrate that centralities inferred by our models are in good agreement with scores computed using the true underlying networks. The learnt mapping produces centralities competitive to that of a two-step method, i.e., a full GP trained on the node-attributes and computed centralities. In addition, we demonstrate how VBC-GP can be used for limited vaccine distribution to constrain the spread of a virus.

2 Background: Graphs and Centrality Measures

We work with networks or graphs that contain vertices or nodes, connected via links or edges. More formally, a graph $G \triangleq (V, E)$ is a tuple consisting of a set of vertices $V = \{v_1, v_2, \ldots, v_{|V|}\}$ and a set of edges $e_k = (v_j, v_i, w_{ij}) \in E$. The variable w_{ij} grants to each edge a weight, which can represent, for example, probabilities of click-throughs in web-surfing or origin-destination counts in transportation modeling.

By labeling each vertex with an integer label, we can associate with graph G a weighted adjacency matrix $\mathbf{W} = [w_{ij}]$; the ordering of the subscripts in w_{ij} indicates an edge from j to i. The networks we deal with can be undirected or directed, with two key assumptions: (i) edge weights are non-negative $w_{ij} \geq 0$ and (ii) G is strongly-connected, i.e., there exist a directed path between any two vertices $v_i, v_j \in V$. The class of graphs that fall under these two assumptions remains broad, e.g., a social network of friends or an interlinked group of web-pages.

2.1 Centrality in Networks

The most significant concept in network analysis is arguably that of node importance or centrality. Since "importance" depends on context, a number of centrality measures (e.g., betweenness, closeness) have been developed and we refer readers to [7] for more details. Let us associate with a graph G a centrality vector $\mathbf{c} = [c_i]_{i=1}^{|V|}$, representing the importance of each node.

Degree Centrality The simplest centrality measure is the degree, i.e., the number of edges incident upon a node. When dealing with directed weighted graphs, vertices can have both an in-degree $c_i^{\text{in}} = \sum_{j \in N(i)} w_{ij}$ and an out-degree $c_i^{\text{out}} = \sum_{j \in N(i)} w_{ji}$, where $N(i)$ is a neighborhood function that returns the neighbors of node i. An practical examples, the "citation-count" metric for estimating scientific paper significance is an in-degree measure and the "influence" of actors in organizational structures is an out-degree measure.

Eigenvector Centrality While conceptually simple, the degree centrality fails to take into account the "value" of linked nodes. It is reasonable that having connections to important nodes increases one's own importance. Let us define the eigenvector centrality of a node to be the weighted sum of the centralities of its neighbors:

$$c_i^{\text{eig}} = \sum_{j \in N(i)} w_{ij} c_j^{\text{eig}} \tag{1}$$

We see that (1) is a generalization of in-degree where in the latter, links are simply assigned a value of one. The centralities are then found by solving the equation:

$$\lambda_1 \mathbf{c}^{\text{eig}} = \mathbf{W} \mathbf{c}^{\text{eig}} \tag{2}$$

where λ_1 is the principal eigenvalue and $\mathbf{c}^{\text{eig}}$ is the associated right eigenvector. There are two notable extensions to the canonical eigenvector centrality: Katz Centrality [8] and Google's PageRank [9]. We hold-off discussion of PageRank until Section 6, where we consider future work.

Katz Centrality One possible issue concerning eqn. (1) is that nodes can possess zero centrality, i.e., those having no incoming edges or nodes only pointed to by the former. The Katz centrality measure advocates a minimal centrality to each node:

$$c_i^{\text{katz}} = a \sum_{j \in N(i)} w_{ij} c_j^{\text{katz}} + b \tag{3}$$

where a and b are positive constants. Typically, $b = 1$, which yields $\mathbf{c}^{\text{katz}} = (\mathbf{I} - a\mathbf{W})^{-1}\mathbf{1}$, where a has become the main parameter of interest. To prevent divergence, a is set below the inverse of the maximal eigenvalue, but is otherwise experimentally tuned. We can adopt the perspective that a and b act as simple "priors" on node centralities. Taking this notion further, we can extend eigenvector centrality in a probabilistic manner to account for link errors, as we discuss below.

3 Variational Bayesian Centrality Gaussian Process (VBC-GP)

Consider a dataset $\mathcal{D} = \{d_k\}$ where each sample is a tuple consisting of a node and its observed incoming edges, $d = \big(i, \{\,(j, \hat{w}_{ij})\,\}_{j \in N(i)}\big)$ where i is the sampled node and $\{(j, \hat{w}_{ij})\}$ is the set of i's observed neighbors and weights. In contrast with existing centrality methods which typically assume a single weight matrix, we allow for repeated observations for a single edge. Using this dataset, we adopt a Bayesian approach for inferring node centralities:

$$p(\mathbf{c}, \lambda | \mathcal{D}) \propto p(\mathcal{D} | \mathbf{c}, \lambda) p(\mathbf{c}) p(\lambda) \tag{4}$$

Starting with the centrality and eigenvalue priors, recall from Section 2 that, by assumption, our underlying network is strongly-connected with non-negative weights. As such, the weighted adjacency matrix induced by our graph is irreducible, allowing us to exercise the Perron-Frobenius theorem [10, 11]:

Theorem 1 (Perron-Frobenius). *Suppose the matrix* $\mathbf{W}$ *is irreducible (the graph is strongly connected), then:*

1. $\mathbf{W}$ *has a positive real eigenvalue* λ_1 *larger (or equal to) in magnitude to all other eigenvalues of* $\mathbf{W}$*,* $\lambda_1 \geq |\lambda|$*.*
2. λ_1 *has algebraic and geometric multiplicity 1.*
3. *Associated with* λ_1 *is a positive eigenvector* $\mathbf{p} > 0$ *(the Perron vector). Any non-negative eigenvector is a multiple of* $\mathbf{p}$*.*

The Perron-Frobenius theorem guarantees that the principal eigenvector $\mathbf{p}$ has all positive elements and λ_1 is a simple eigenvalue with a single *unique* eigenvector (excluding positive multiples). In other words, to compute the centralities, we have to ensure $\mathbf{c}$ is positive. We place log-normal priors, with open support $(0, \infty)$, on $\mathbf{c}$ and λ to enforce this constraint and incorporate prior notions of node centralities. To facilitate later connections to Gaussian processes (GP), we parametrize our model using Gaussian random variables $\mathbf{z} = [z_i]$ where $c_i = \exp(z_i)$[1]. Likewise, $\lambda = \exp(z_\lambda)$, yielding priors with the form:

$$p(\mathbf{c}) = \prod_i^{|V|} \exp(z_i) \quad \text{and} \quad p(\lambda) = \exp(z_\lambda) \tag{5}$$

where

$$p(z_i) = \mathcal{N}(\mathring{\mu}_i^2, \mathring{\sigma}_i^2) \quad \text{and} \quad p(z_\lambda) = \mathcal{N}(\mathring{\mu}_\lambda^2, \mathring{\sigma}_\lambda^2) \tag{6}$$

The choice of likelihood function $p(\mathcal{D}|\mathbf{z}, z_\lambda)$ will differ depending on application and the underlying observation process. Here, we derive a variational bound based on the assumption that the difference between the log latent and computed centralities are normally distributed, which was found to be

[1]The exponentiation of a normal r.v. $z \sim \mathcal{N}(\mu, \sigma^2)$ yields a log-normal r.v., $\exp(z_i) \sim \ln\mathcal{N}(\mu, \sigma^2)$.

reasonable under varying edge noise conditions. Let $i = i(k)$ denote the k-th observed node, then,

$$
\begin{aligned}
p(\mathcal{D}|\mathbf{z}, z_\lambda) &= \prod_{d \in \mathcal{D}} p(d_k|\mathbf{z}, z_\lambda) \\
p(d_k|\mathbf{z}, z_\lambda) &= \frac{1}{\sqrt{2\pi\tilde{\sigma}_i^2}} \exp\left\{ -\frac{(\log(\sum_{j \in N(i)} \hat{w}_{ij} \exp(z_j)) - \log(\exp(z_\lambda)\exp(z_i)))^2}{2\tilde{\sigma}_i^2} \right\} \\
&= \frac{1}{\sqrt{2\pi\tilde{\sigma}_i^2}} \exp\left\{ -\frac{(\zeta_i - z_\lambda - z_i)^2}{2\tilde{\sigma}_i^2} \right\} \qquad (7)
\end{aligned}
$$

where $\zeta_i = \log\big(\sum_{j \in N(i)} \hat{w}_{ij} \exp(z_j)\big)$. To simplify our exposition, we point-optimize the node-dependent noise terms $\tilde{\sigma}_i$ but a prior can be adopted if necessary. Other likelihoods can be applied to induce different centrality measures and fit within the preceding framework.

3.1 Mapping Centralities to Node Attributes

Now, consider that each node i is associated with an observed feature vector $\mathbf{x}_i$ and the centralities $\mathbf{c}$ (and hence, $\mathbf{z}$) can be well-modelled by a latent function on the node attributes $f(\mathbf{x})$. In this work, we use a variational sparse GP [12, 13] and introduce the the auxiliary inducing variables $\mathbf{u}$, the inducing input locations $\hat{\mathbf{X}} = [\hat{\mathbf{x}}_i]_{i=1}^m$ and the distributions,

$$
\begin{aligned}
p(\mathbf{z}|\mathbf{u}, \hat{\mathbf{X}}) &= \mathcal{N}(\mathbf{K}_{nm}\mathbf{K}_{mm}^{-1}\mathbf{u}\,,\, \mathbf{K}_{nn} - \mathbf{K}_{nm}\mathbf{K}_{mm}^{-1}\mathbf{K}_{mn}) \\
p(\mathbf{u}) &= \mathcal{N}(\mathbf{0}, \mathbf{K}_{mm})
\end{aligned}
$$

where $k_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$, $\mathbf{k}_i = [k(\mathbf{x}_i, \hat{\mathbf{x}}_j)]_{j=1}^m$ and $\mathbf{K}_{mm} = [k(\hat{\mathbf{x}}_i, \hat{\mathbf{x}}_j)]_{i,j=1}^m$. Taking a step-back, we see now that,

$$
p(\mathbf{u}, z_\lambda|\mathcal{D}) \propto p(\mathcal{D}|\mathbf{z}, \lambda) p(\mathbf{z}|\mathbf{u}, \hat{\mathbf{X}}) p(\mathbf{u}) p(z_\lambda) \qquad (8)
$$

where $\mathbf{z}$ is now represented indirectly via the sparse GP and the inducing variables $\mathbf{u}$. This variant enables the prediction of centralities from node features and permits a degree of compression—centralities for a potentially large number of nodes can be represented using a small number of inducing variables.

3.2 An Approximate Variational Lower Bound

Our goal now is to infer the latent $\mathbf{u}$ inducing variables (and z_λ), which can be achieved with MCMC or via a variational distribution. The latter is generally more computationally efficient and is the approach undertaken in this paper. We assume a fully-factorized mean-field variational distribution,

$$
\begin{aligned}
q(\mathbf{z}, \mathbf{u}, z_\lambda) &= p(\mathbf{z}|\mathbf{u}, \hat{\mathbf{X}}) q(\mathbf{u}) q(z_\lambda) \\
q(\mathbf{u}) &= \mathcal{N}(\mathbf{m}, \mathbf{S}) \\
q(z_\lambda) &= \mathcal{N}(\mu_\lambda^2, \sigma_\lambda^2) \qquad (9)
\end{aligned}
$$

and formulate a lower bound,

$$
\begin{aligned}
\mathcal{L}_1(q) &= \iiint p(\mathbf{z}|\mathbf{u}, \hat{\mathbf{X}}) q(\mathbf{u}, z_\lambda) \log \frac{p(\mathcal{D}|\mathbf{z}, \lambda) p(\mathbf{u}) p(z_\lambda)}{q(\mathbf{u}, z_\lambda)} d\mathbf{z} d\mathbf{u} dz_\lambda \\
&= \mathbb{E}[\log p(\mathcal{D}|\mathbf{z}, \lambda)] - \mathbb{D}_{\mathrm{KL}}[q(\mathbf{u}\|p(\mathbf{u})] - \mathbb{D}_{\mathrm{KL}}[q(z_\lambda)\|p(z_\lambda)]. \qquad (10)
\end{aligned}
$$

where the term $p(\mathbf{z}|\mathbf{u}, \hat{\mathbf{X}})$ inside the logarithm cancels out. We can find $q(\mathbf{u}, z_\lambda)$ by maximizing $\mathcal{L}_1$ (or equivalently, minimizing $-\mathcal{L}_1$). The last two terms on the RHS are the KL-divergences between two normal distributions and are easily computed,

$$
\mathbb{D}_{\mathrm{KL}}(\mathcal{N}_q\|\mathcal{N}_p) = \frac{1}{2}\left(\mathrm{tr}\left(\boldsymbol{\Sigma}_p^{-1}\boldsymbol{\Sigma}_q\right) + \left(\boldsymbol{\mu}_p - \boldsymbol{\mu}_q\right)^{\mathrm{T}} \boldsymbol{\Sigma}_p^{-1}(\boldsymbol{\mu}_p - \boldsymbol{\mu}_q) - d - \ln \frac{|\boldsymbol{\Sigma}_q|}{|\boldsymbol{\Sigma}_p|} \right). \qquad (11)
$$

The principal difficulty is computing the expectation $\mathbb{E}[\log p(\mathcal{D}|\mathbf{z}, \lambda)]$, which is taken with respect to $q(\mathbf{u})$ and $p(\mathbf{z}|\mathbf{u}, \hat{\mathbf{X}})$:

$$
\mathbb{E}[\log p(\mathcal{D}|\mathbf{z}, \lambda)] = -\frac{1}{2}\sum_{k=1}^{|\mathcal{D}|} \log 2\pi\tilde{\sigma}_i^2 - \sum_{k=1}^{|\mathcal{D}|} \frac{\mathbb{E}[(\zeta_i - z_\lambda - z_i)^2]}{2\tilde{\sigma}_i^2}. \qquad (12)
$$

Since the remaining expectation, $\mathbb{E}[(\zeta_i - z_\lambda - z_i)^2]$, does not admit a simple closed-form solution, we apply Taylor approximations,

$$\begin{aligned}\mathbb{E}[(\zeta_i - z_\lambda - z_i)^2] &\approx (\mathbb{E}[\zeta_i - z_\lambda - z_i])^2 \\ &= (\mathbb{E}[\zeta_i] - \mathbb{E}[z_\lambda] - \mathbb{E}[z_i])^2 \\ &= (\mathbb{E}[\zeta_i] - \mu_\lambda - \mathbf{k}_i^\top \mathbf{K}_{mm}^{-1}\mathbf{m})^2 \\ &\approx (\hat{\xi}_i - \mu_\lambda - \mathbf{k}_i^\top \mathbf{K}_{mm}^{-1}\mathbf{m})^2 \end{aligned} \tag{13}$$

where, for the final step, we have exploited the linearity of expectations and the assumed independence between the random variable to derive the approximation $\hat{\xi}_i \approx \mathbb{E}[\zeta_i]$:

$$\hat{\xi}_i = \log\left(\sum_{j\in N(i)} \hat{w}_{ij}\mathbb{E}[\exp(z_j)]\right) - \frac{\sum_{j\in N(i)} \hat{w}_{ij}^2 \mathbb{V}[\exp(z_j)]}{2(\sum_{j\in N(i)} \hat{w}_{ij}\mathbb{E}[\exp(z_j)])^2} \tag{14}$$

The expectation and variance of $\exp(z_j)$ are given by

$$\mathbb{E}[\exp(z_j)] = \exp(\mathbf{k}_j^\top \mathbf{K}_{mm}^{-1}\mathbf{m} + \hat{\sigma}_j^2/2 + \mathbf{b}^\top \mathbf{S}\mathbf{b}/2) \tag{15}$$

$$\mathbb{V}[\exp(z_j)] = (\exp(\mathbf{b}^\top \mathbf{S}\mathbf{b}) - 1)\exp(2\mathbf{k}_j^\top \mathbf{K}_{mm}^{-1}\mathbf{m} + \hat{\sigma}_j^2 + \mathbf{b}^\top \mathbf{S}\mathbf{b}) \tag{16}$$

where $\hat{\sigma}_j = k_{jj} - \mathbf{k}_j^\top \mathbf{K}_{mm}^{-1}\mathbf{k}_j$ (the predictive variance) and $\mathbf{b} = (\mathbf{k}_i^\top \mathbf{K}_{mm}^{-1})^\top$. In the derivation of the expectation and variance above, we have used the fact that the transformation $\mathbf{k}_j^\top \mathbf{K}_{mm}^{-1}\mathbf{u} + \hat{\sigma}_j^2/2$ results in a normal distribution $\mathcal{N}(\mathbf{k}_j^\top \mathbf{K}_{mm}^{-1}\mathbf{m} + \hat{\sigma}_j/2, \mathbf{b}^\top \mathbf{S}\mathbf{b})$.

Summarized Lower-bound: Putting all the terms together results in the approximate lower bound:

$$\begin{aligned}\mathcal{L}_2 = &-\frac{1}{2}\sum_{k=1}^{|\mathcal{D}|}\log 2\pi\tilde{\sigma}_i^2 - \sum_{k=1}^{|\mathcal{D}|}\frac{1}{2\tilde{\sigma}_i^2}\left[(\hat{\xi}_i - \mu_\lambda - \mathbf{k}_i^\top \mathbf{K}_{mm}^{-1}\mathbf{m}\Big)^2\right] \\ &- \mathbb{D}_{\mathrm{KL}}(q(\mathbf{u})\|p(\mathbf{u})) - \mathbb{D}_{\mathrm{KL}}(q(z_\lambda)\|p(z_\lambda)).\end{aligned} \tag{17}$$

Maximizing $\mathcal{L}_2$ yields the approximate posterior $q(\mathbf{u})q(z_\lambda)$, which can be used to obtain the centrality of a node $\mathbf{x}_*$—we simply exponentiate the normal predictive distribution, which results in a log-normal $c_* \sim \ln\mathcal{N}(\mathbf{k}_*^\top \mathbf{K}_{mm}^{-1}\mathbf{m}, k_{**} - \mathbf{k}_*^\top \mathbf{K}_{mm}^{-1}\mathbf{k}_* + \mathbf{b}^\top \mathbf{S}\mathbf{b})$. In this work, we have used conjugate-gradient optimization, but given large datasets, $\mathcal{L}_2$ can be maximized using mini-batches or stochastic gradient ascent.

4 Experimental Results

In this section, we present results of experiments using VBC-GP to infer centrality on three real-world networks, i.e., a social network of a Facebook ego user (FB) [14], Freeman's network of communication between 32 academics (FM) [15] and the Singapore public-transportation people-flow network (PT).

Table 1: Network Dataset Properties. Number of nodes $|V|$, Number of edges $|E|$, node attributes, weights, average (unweighted) degree $\langle k\rangle$ and average (unweighted) clustering coefficient $\langle C_c\rangle$.

	$\lvert V\rvert$	$\lvert E\rvert$	Attr.	Weights	$\langle k\rangle$	$\langle C_c\rangle$
FM	32	552	3	Y	6.88	0.86
FB	224	6384	3	N	28.50	0.54
PT	57	3249	2	Y	0.94	1.00

Basic properties of the networks are given in Table 1. The node attributes for FM were the total correspondence, department and total number of citations received. For FB, nodes were associated with the number of contacts and anonymized binary features (reduced to two dimensions via PCA). Finally, for PT, the attributes were the total number of travellers for each location and the resident population. For each network, we corrupted the links weights with log-normal noise having variance $\sigma_o^2 = 5.0$.

Our models, datasets and experiment code are available[2] and can be used to reproduce the following results.

[2]On the author's website at http://www.haroldsoh.com.

4.1 Mapping Centrality with Noisy Edges

We provided VBC-GP with an increasing number of samples per node ($N_s = 1, 2, \ldots, 10$) and each "run" was repeated 15 times. The lower-bounds were optimized using a conjugate-gradients method with the initial point set using the eigenvector centrality of an averaged weight matrix $\overline{\mathbf{W}} = \left[N_{ij}^{-1} \sum \hat{w}_{ij}\right]_{i,j=1}^{|V|}$ where N_{ij} is the number of observations for weight w_{ij}.

The VBC-GP was set to use a squared exponential kernel with an initial lengthscale of $l = 15, 20, 20$ with $m = 10, 20, 40$ inducing inputs for FM, PT and FB respectively. Following [16], the inducing input locations $\hat{\mathbf{X}}$ were selected via k-means clustering. To simplify the optimization process, we parameter-tied the noise terms $\tilde{\sigma}_i^2 = \tilde{\sigma}_n^2$.

As a performance measure, we use the Kendall correlation to eigenvector centralities of the noise-free network[3]. As a baseline model, we trained a full GP with hyperparameter optimization (no sparsity) on the averaged weight centralities using 80% of the nodes. The GP was then used to predict the centralities of all the nodes, which were compared to the true centralities via Kendall correlation.

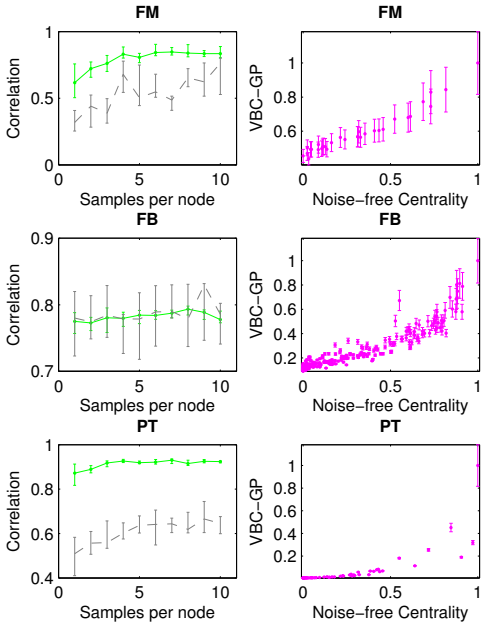

Figure 1: The VBC-GP learned high-quality mappings compared to the full-GP (dashed-line) with higher correlation scores as N_s increased (FM and PT). A decrease in the interquartile ranges suggests more robust performance. Scatter plots (on the right) show the predicted centralities.

The scores obtained by full-GP (dashed line) and VBC-GP are shown in Fig. 1. Similar to the VBC, the VBC-GP obtained better scores with more samples and outperformed the full-GP on the FM and PT networks (higher correlation scores of 7.1% and 27.8% at $N_s = 10$ respectively). The averaging method appeared to slightly outperform the VBC-GP on FB (0.3% at $N_s = 10$). With all three networks, the interquartile ranges for the VBC-GP results were narrower—a tell-tale sign of more robust performance. Overall, the VBC-GP predicted centralities—using only 10 to 40 inducing inputs—were strongly correlated with the true centralities for all three networks, as can be visually confirmed by the sample scatter plots (right side of Fig. 1).

5 Case Study: Vaccine Distribution

How to distribute a limited number of vaccines or prophylactic treatment in a susceptible population remains a difficult and relevant problem in epidemiology. In this case study, we used VBC-GP to identify central individuals to immunize, with the aim of preventing/slowing disease spread. We simulated viral disease outbreaks on contact networks using discrete-time Susceptible-Infectious-Recovered (SIR) dynamics, i.e., infected agents infect susceptibles who, in turn, may spread the disease before recovering. The disease had a time-independent probability of transmission, $p(I_t) = 0.05$, with recovery rate 0.1.

The contact network was generated using location-based hierarchical link formation [17]; for example, other than family members, students are assumed to form contacts primarily in classrooms and to a lesser extent, with other students in the same school. Likewise, contacts are probabilistically sampled for workplaces (firms comprising departments), pre-schools, universities and residential communities. The edge weights, $w_{ij} \in [0, 1]$ represent the probability of a disease transmitting contact. As such, the virus propagates along a given edge e_{ij} with probability $w_{ij}p(I_t)$. The observed

[3]We also computed the Pearson and Spearman correlation and top-ten overlap score. All the scores were strongly-correlated and hence, only the Kendall score is shown here for conciseness.

weights are noisy samples from a truncated normal $\mathcal{N}(w_{ij}, 0.5)$ in the interval $[0, 1]$. Each individual is associated with node attributes comprising his/her age and the location-dependent sizes (which were reduced to 3 dimensions using PCA). Note that the location attributes are *not* the number of contacts formed, e.g., a student may form only 5 contacts in a class-size of 10.

Fig. 2 (top) illustrates the epidemic profile—averaged over 100 simulation runs on 15 different networks starting with 1% infected—comparing scenarios with zero vaccination, random baseline, Full-GP baseline and sparse VBC-GP (40 inducing inputs) distribution of vaccines in a population of 360 people (12,000 edges, 30% vaccination). Treating the VBC-GP selected individuals decreased the total number of infected by 62% (239.5 to 90.7 people) and the peak (91.0 to 20.4 people and a delay from 23 to 33 days)—better than the baselines.

To evaluate model generalization, we constructed a larger population of $\approx 17,000$ people and used a VBC-GP model (trained on the smaller networks) to select 30% of the nodes for vaccination, using only *predicted* node centralities. Fig. 2 (bottom) shows that the VBC-GP-based vaccine distribution on the novel network (sampled using the same process) resulted in a similar positive outcomes, i.e., a decrease in the total infected by 41% (from 9983 to 5853 people) and the peak (4319 to 1686 people).

In summary, immunizing the VBC-GP selected nodes was superior to the alternative methods in reducing the severity of the outbreaks—a sharp fall in the peak and total infected reduces the strain on healthcare systems. Although our study focusses on a biological virus, similar analyses can be conducted for mitigating the spread of computer viruses or reducing the impact of disruptions/attacks on transportation hubs.

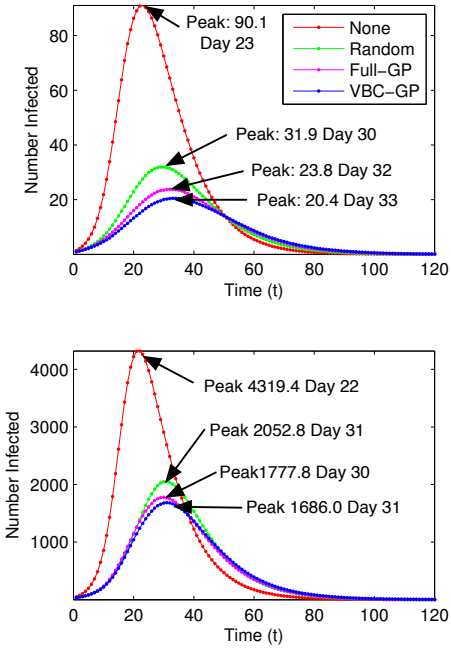

Figure 2: Disease progression profiles in terms of number of infected over time (averaged over 100 simulation). (top) Network of $\approx$ 360 people (12,000 edges) and (bottom) Larger network of $\approx$ 17,000 people (540,000 edges). In both cases, vaccinating the VBC-GP selected nodes resulted in a drop in the total and peak number of infected.

6 Conclusions and Future Work

In this paper, we contributed the VBC-GP—a Bayesian model of eigenvector centrality that uses a sparse GP to map node-attributes to latent centralities. The VBC-GP alleviates the need to maintain separate distributions for each node and permits prediction of centralities directly from node-attributes (as shown in the case-study). The results show that high-quality mappings could be learned in the presence of noise. The correlations compare favorably to the full-GP, with additional benefits: the model is sparse and the centralities are computed along with the mapping in a single optimization process.

We anticipate that the proposed model can be improved in several ways. Regarding scalability, real-world networks can be very large, comprising millions of nodes. The VBC-GP—promising given its sparse nature—may be trained using the natural gradient [16].

Although the Gaussian likelihood worked well in our experiments, it may be inappropriate where the noise process results in different (perhaps multi-modal) deviation distributions. In such cases, an alternative likelihood function can be substituted into (4). Different likelihoods can also be applied to induce different notions of centrality, e.g., PageRank is a variant of eigenvector centrality with centralities weighted by the outgoing links. Incorporating such a weighting into the likelihood model would result in a Bayesian form of PageRank centrality. Experimentation with different likelihoods under varying noise conditions would make interesting future work.

Finally, taking a broader perspective, this paper bridges metrics for complex networks with modern machine learning and we hope that it will catalyze development of future network metric models.

References

[1] J. Platig, E. Ott, and M. Girvan, "Robustness of network measures to link errors," *Phys. Rev. E*, vol. 88, p. 062812, Dec 2013.

[2] T. L. Frantz, M. Cataldo, and K. M. Carley, "Robustness of centrality measures under uncertainty: Examining the role of network topology," vol. 15, no. 4, pp. 303–328, 2009.

[3] S. P. Borgatti, K. M. Carley, and D. Krackhardt, "On the robustness of centrality measures under conditions of imperfect data," *Social networks*, vol. 28, no. 2, pp. 124–136, 2006.

[4] E. Bullmore and O. Sporns, "The economy of brain network organization," *Nature Reviews Neuroscience*, vol. 13, no. 5, pp. 336–349, 2012.

[5] M. Kim and J. Leskovec, "Multiplicative attribute graph model of real-world networks," *Internet Mathematics*, vol. 8, no. 1-2, pp. 113–160, 2012.

[6] K. Palla, D. A. Knowles, and Z. Ghahramani, "An infinite latent attribute model for network data," in *Proceedings of the 29th International Conference on Machine Learning*, July 2012.

[7] M. Newman, *Networks: An Introduction*. Oxford University Press, 2010.

[8] L. Katz, "A new status index derived from sociometric analysis," *Psychometrika*, vol. 18, no. 1, pp. 39–43, 1953.

[9] S. Brin and L. Page, "The anatomy of a large-scale hypertextual web search engine," *Computer networks and ISDN systems*, vol. 30, no. 1, pp. 107–117, 1998.

[10] O. Perron, "Zur theorie der matrices," vol. 64, no. 2, pp. 248–263, 1907.

[11] G. Frobenius, "Uber Matrizen aus nicht negativen Elementen," pp. 456–477, 1912.

[12] M. Titsias, "Variational Learning of Inducing Variables in Sparse Gaussian Processes," in *The 12th International Conference on Artificial Intelligence and Statistics (AISTATS)*, vol. 5, 2009.

[13] J. Quiñonero Candela and C. E. Rasmussen, "A unifying view of sparse approximate gaussian process regression," *Journal of Machine Learning Research*, vol. 6, pp. 1939–1959, 2005.

[14] J. J. McAuley and J. Leskovec, "Learning to discover social circles in ego networks.," in *NIPS*, vol. 272, pp. 548–556, 2012.

[15] S. C. Freeman and L. C. Freeman, *The networkers network: a study of the impact of a new communications medium on sociometric structure*. Sch. of Soc. Sciences, Uni. of Calif., 1979.

[16] J. Hensman, N. Fusi, and N. D. Lawrence, "Gaussian Processes for Big Data," in *Uncertainty in Artificial Intelligence (UAI-13)*, 2013.

[17] T. Zhang, X. Fu, C. K. Kwoh, G. Xiao, L. Wong, S. Ma, H. Soh, G. K. K. Lee, T. Hung, and M. Lees, "Temporal factors in school closure policy for mitigating the spread of influenza," *Journal of public health policy*, vol. 32, no. 2, pp. 180–197, 2011.