Variational Bayesian Network Centrality

Harold Soh
Singapore-MIT Alliance for Research and Technology (SMART)
Singapore 138602

September 16, 2014

Abstract

We present an eigenvector-based Bayesian centrality model for determining node importance in a graph or network. In contrast to existing methods, our model explicitly considers noisy weighted links and thus, allows for the assimilation of multiple observations, the inclusion of priors and the extraction of uncertainties. To enable tractable inference, we develop a variational lower bound that is demonstrated to be effective on a variety of networks (two synthetic and six real-world graphs). We further extend our approach incorporate node attributes, yielding the sparse variational Bayesian centrality Gaussian process (VBC-GP). This model learns a mapping between node attributes/features to latent centrality and hence, is able to represent a large number of nodes using only a limited number of inducing inputs. Experiments show that the VBC-GP learns high-quality mappings and compares favorably to a full-GP trained directly on the node attributes and true network centralities.

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 developed over the years. Among the variety of analytical instruments proposed, node importance or centrality has potentially found the most widespread significance. To give a few examples, centrality measures were used in the identification of gene-disease associations [1], important actors in a social network [2], city travel-hubs [3], and relevant pages in web search [4].

However, 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 [5,6,7]. In particular, the eigenvector
centrality measure undergoes a smooth decay in accuracy (relative to the centrality of the true network) as link errors increase \[7\]. This issue is important in both theoretical and applied settings, e.g., in neuroscience, centrality scores give us insights into brain organization \[8\] and erroneous scores may falsely identify hub modules.

**Summary of Contributions** The contributions in this work are twofold. First, we develop a novel Bayesian centrality model for networks with noisy weighted links. Our model extends the popular eigenvector-based measure by considering centralities as latent attributes that are constrained by the eigenvector centrality relation. By treating centrality estimation as probabilistic latent variable inference, this approach directly addresses the problem of uncertainty when inferring node importance and permits principled assimilation of repeated weight observations. Moreover, the model allows for the incorporation of prior knowledge regarding the centralities and the extraction of uncertainties via the posterior. To enable tractable inference, we present a variational lower-bound that can be optimized either offline or iteratively.

Second, we extend the variational model to include information from node attributes via a sparse Gaussian Process (GP) \[9, 10\]. In effect, our model 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 seven different datasets (two synthetic and six real-world networks) demonstrate that centralities inferred by our models are in good agreement with scores computed using the true underlying networks. In addition, the learnt mapping produces centralities competitive to that of a two-step method, i.e., a full GP trained on the node-attributes and true centralities.

2 Background: Graphs and Centrality Measures

In this paper, we work with networks or graphs that contain vertices or nodes, connected via links or edges. More formally, a graph $G \doteq (V, E)$ 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 $W = [w_{ij}]$; 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: (1) edge weights are non-negative $w_{ij} \geq 0$ and (2) $G$ 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 webpages.

2.1 Centrality in Networks

As mentioned in the introduction, 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 have been developed; we refer readers to [11] for more details. Let us associate with a graph $G$ a centrality vector $c = [c_i]_{i=1}^{|V|}$, 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 $c^\text{in}_i = \sum_{j \in N(i)} w_{ij}$ and an out-degree $c^\text{out}_i = \sum_{j \in N(i)} w_{ji}$, 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:

$$c^\text{eig}_i = \sum_{j \in N(i)} w_{ij} c^\text{eig}_j$$

(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:

$$\lambda_1 c^\text{eig} = W c^\text{eig}$$

(2)

where $\lambda_1$ is the principal eigenvalue and $c^\text{eig}$ is the associated right eigenvector.

There are two notable extensions to the canonical eigenvector centrality: Katz Centrality [12] and Google’s PageRank [4]. 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:

$$c^\text{katz}_i = a \sum_{j \in N(i)} w_{ij} c^\text{katz}_j + b$$

(3)
where \( a \) and \( b \) are positive constants. Typically, \( b = 1 \), which yields:

\[
c^{\text{katz}} = (I - aW)^{-1}1
\]  

(4)

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 Bayesian Node Centrality

Consider a dataset \( D = \{d_k\} \) where each sample is a tuple consisting of a node and its observed incoming edges, \( d = (i, \{(j, \hat{w}_{ij})\}_{j \in N(i)}) \) 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:

\[
p(c, \lambda |D) \propto p(D|c, \lambda)p(c)p(\lambda)
\]  

(5)

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 [13, 14]:

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

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

The Perron-Frobenius theorem guarantees that the principal eigenvector \( p \) has all positive elements. Furthermore, \( \lambda_1 \) is a simple eigenvalue with a single unique eigenvector (excluding positive multiples). In other words, to compute the centralities, we have to ensure \( c \) is positive. We place flexible Gamma priors, with open support \((0, \infty)\), on \( c \) and \( \lambda \) to enforce this constraint and incorporate prior notions of node importances:

\[
p(c) = \prod_i \text{Ga}(\alpha_0^i, \beta_0^i) \quad \text{and} \quad p(\lambda) = \text{Ga}(\alpha_0^\lambda, \beta_0^\lambda)
\]

The choice of likelihood function \( p(D|c, \lambda) \) will differ depending on application and the underlying observation process. As a first step, we demonstrate a variational bound based on the assumption that the difference between the
latent and computed centralities (given the observed weights) are approximately zero-mean normal,

\[
\log p(D|c, \lambda) = \sum_{d \in D} \log \mathcal{N}\left(0 | \lambda c_i - \sum_{j \in \mathcal{N}(i)} \hat{w}_{ij} c_j, \sigma_n^2 \right)
\]

which we found to be reasonable under different edge noise conditions (See Supplementary Material for empirical fits under normal, log-normal and Poisson noise). As we will see in Section 5, the Gaussian likelihood performs well even under relatively high noise conditions that delete a proportion of edges. Other likelihoods (e.g., the truncated or log-normal) can be more appropriate under different circumstances and fit within the preceding framework. To simplify the following derivations, a single noise variance \(\sigma_n^2\) term is used; we discuss reintroducing multiple noise terms in Section 6. In the next section, we derive a variational lower-bound for maximizing \(\log p(D)\), which will allow us to infer the model posterior.

### 3.1 Variational Lower Bound

We adopt a mean-field variational approximation using a fully-factorized posterior:

\[
q(c, \lambda) = \prod_i q(c_i) q(\lambda), \quad q(c_i) = \mathcal{N}(\mu_i, \sigma_i^2), \quad q(\lambda) = \text{Ga}(\alpha_\lambda, \beta_\lambda)
\]

Using Gaussian approximations for the centralities facilitates our later connection to node attributes via a Gaussian process (in Section 4), but alternative distributions (e.g., Gamma) can be used. With this factorization in mind, we write the variational lower-bound:

\[
\mathcal{L}_1(q) = \int \int q(c, \lambda) \log \frac{p(D|c, \lambda)p(c)p(\lambda)}{q(c, \lambda)} dcd\lambda
\]

\[
= \mathbb{E}[\log p(D|c, \lambda)] + \mathbb{E}[\log p(c)] - \mathbb{E}[\log q(c)] + \mathbb{E}[\log p(\lambda)] - \mathbb{E}[\log q(\lambda)]
\]

We demonstrate that each of these expectations can be analytically obtained or Taylor-approximated. To simplify notation let \(i = i(d)\), i.e., the observed node in sample \(d\).

**Computing \(\mathbb{E}[\log p(D|c, \lambda)]\):** Given the Gaussian likelihood, this expectation is given by:

\[
\mathbb{E}[\log p(D|c, \lambda)] = -\frac{|D|}{2} \log \sigma_n^2 - \frac{1}{2\sigma_n^2} \sum_{d \in D} \zeta_d + \text{const} \quad (7)
\]
where $\zeta_d$ is the square within the normal distribution with the expectations brought in:

$$
\zeta_d = \mathbb{E}\left[\left(\lambda c_i - \sum_{j \in N(i)} \hat{w}_{ij} c_j\right)^2\right] = \mathbb{E}[\lambda^2 c_i^2] - 2\mathbb{E}[\lambda c_i] + \mathbb{E}[\hat{c}_i^2]
$$

where we have defined $\hat{c}_i = \sum_{j \in N(i)} \hat{w}_{ij} c_j$, which has mean, $\hat{\mu}_i = \sum_{j \in N(i)} \hat{w}_{ij} \mu_j$, and variance, $\hat{\sigma}_i^2 = \sum_{j \in N(i)} \hat{w}_{ij}^2 \sigma_j^2$. Given our assumed factorized distribution, we can separate out the terms:

$$
\zeta_d = \mathbb{E}[\lambda^2] \mathbb{E}[c_i^2] - 2\mathbb{E}[\lambda] \mathbb{E}[c_i] \mathbb{E}[\hat{c}_i] + \mathbb{E}[\hat{c}_i^2]
$$

and each expectation is easily computed:

$$
\mathbb{E}[c_i] = \mu_i \quad \mathbb{E}[c_i^2] = \mu_i^2 + \sigma_i^2 \quad \mathbb{E}[\hat{c}_i] = \hat{\mu}_i \quad \mathbb{E}[\hat{c}_i^2] = \hat{\mu}_i^2 + \hat{\sigma}_i^2
$$

**Approximating $\mathbb{E}[\log p(c)]$:** We give a first-order Taylor approximation of the expectation of the Gamma likelihood under a Gaussian random variable:

$$
\mathbb{E}[\log p(c)] = \sum_{i=1}^{V} \left[ (\alpha_i^0 - 1)(\mathbb{E}[c_i]) - \beta_i \mathbb{E}[c_i] \right] + \text{const}
$$

$$
\approx \sum_{i=1}^{V} \left[ (\alpha_i^0 - 1)(\log \mathbb{E}[c_i]) - \beta_i \mathbb{E}[c_i] \right] = \sum_{i=1}^{V} \left[ (\alpha_i^0 - 1)(\log \mu_i) - \beta_i \mu_i \right].
$$

**Computing the remaining terms:** We can recognise $\mathbb{E}[\log q(c)]$ as the entropy of a multivariate Gaussian with diagonal covariance matrix $\Sigma_c$ with variances $\sigma_i^2$ along the diagonal,

$$
\mathbb{H}(c) = \frac{|V|}{2} (1 + \log 2\pi) + \frac{1}{2} \log |\Sigma_c| \quad (11)
$$

and $\mathbb{E}[\log p(\lambda)] - \mathbb{E}[\log q(\lambda)]$ as the negative KL-divergence between the approximated Gamma posterior and the prior:

$$
\mathbb{D}_{KL}(q(\lambda)||p(\lambda)) = (\alpha_\lambda - \alpha_\lambda^0) \psi(\alpha_\lambda) - \log \Gamma(\alpha_\lambda) + \alpha_\lambda^0 \log \beta_\lambda - \alpha_\lambda^0 \log \beta_\lambda^0 + \frac{\alpha_\lambda(\beta_\lambda^0 - \beta_\lambda)}{\beta_\lambda} + \log \Gamma(\alpha_\lambda^0) \quad (12)
$$
Summarized Bound: Finally, combining the terms (and collecting the constants w.r.t. the terms being optimized) yields a new lower bound $L_2$:

$$L_2(q) = -\frac{|D|}{2} \log \sigma_n^2 - \frac{1}{2\sigma_n^2} \sum_{d \in D} \zeta_d + \sum_{i=1}^{|V|} \left[ (\alpha^0_i - 1)(\log \mu_i) - \beta_i \mu_i \right] - H(c) - D_{KL}(q(\lambda)||p(\lambda)) + \text{const} \quad (13)$$

We can find $q(c, \lambda)$ by maximizing $L_2$ (or equivalently, minimizing $-L_2$). In this work, we have used an interior-point method, but given large datasets, $L_2$ can be maximized using mini-batches or stochastic gradient ascent.

4 A Link to Node Attributes via Sparse Gaussian Processes

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 [15] connects nodes probabilistically based on binary attributes and affinity matrices. In [16], Palla, Knowles and Ghahramani propose a hierarchical Bayesian model using latent node feature vectors to capture the probability of links.

Inspired by this approach, we introduce the variational Bayesian centrality Gaussian Process (VBC-GP), which extends our previous approximation to consider node attributes as determining factors of centrality in a network. Consider that each node $i$ is associated with an observed feature vector $x_i$ and the centralities $c$ can be well-modelled by a latent function on the node attributes $f(x)$. In this work, we adopt a technique based on sparse GPs using inducing inputs [9, 10], yielding a new variational distribution:

$$q(c, \lambda|X) = q(\lambda) \int q(u)p(c|u, X)du \quad (14)$$

where we introduce the auxiliary inducing variables $u$, inducing inputs $z_i \in Z$, and the distributions,

$$p(c|u, X) = N(K_{nn}^{-1}u, K_{nn} - K_{nm}K_{mm}^{-1}K_{mn})$$
$$q(u) = N(m, S)$$

where $k_{ij} = k(x_i, x_j)$, $k_i = [k(x_i, z_j)]_{j=1}^m$ and $K_{mm} = [k(z_i, z_j)]_{i,j=1}^m$. In a slightly unconventional way, the sparse GP makes its way into the approximation but not into the original model; although this restricts the mapping to the variational distribution, it allows us to retain our simple Gamma priors over $c$.
and translates into a new lower bound:

\[
\mathcal{L}_d = \int\int q(c, \lambda) \log \frac{p(D|c, \lambda)p(c)p(\lambda)}{q(\lambda) \int q(u)p(c|u, X)du} dcd\lambda
\]

\[
= \mathbb{E}[\log p(D|c, \lambda)] + \mathbb{E}[\log p(c)] - \mathbb{E} \left[ \log \int q(u)p(c|u, X) \right] - \mathbb{D}_{KL}(q(\lambda)||p(\lambda))
\]

(15)

From (15), we see that our approximations for \( \lambda \) and \( c \) remain intact, allowing us to re-use the derivations developed in Section 4. In the following, we drop the explicit dependence on \( X \) and \( Z \) for notational convenience.

**Computing** \( \mathbb{E}[\log p(D|c, \lambda)] \):

Rearranging terms and plugging in eqn. (7):

\[
\mathbb{E}[\log p(D|c, \lambda)] = -\frac{|D|}{2} \log \sigma_n^2 - \frac{1}{2\sigma_n^2} \sum_{d \in D} \int q(u) \zeta_d du + \text{const}
\]

(16)

The main computation involved lies in computing the expectation of \( \zeta_d \) (8):

\[
\Lambda_d = \int q(u) \zeta_d du = \frac{\alpha_\lambda(\alpha_\lambda + 1)}{\beta_\lambda^2} \left( \text{Tr}(\hat{A}_{ii}S) + m^T \hat{A}_{ii}m + \sigma_i^2 \right) - \frac{2\alpha_\lambda}{\beta_\lambda} \sum_{j \in N(i)} \tilde{w}_{ij} \left( \text{Tr}(\hat{A}_{ij}S) + m^T \hat{A}_{ij}m \right) + \text{Tr}(\hat{B}_dS) + m^T \hat{B}_d m + \tilde{\sigma}_i
\]

(17)

where \( \hat{A}_{ij} = K_{i,m}^{-1}k_j^T K_{m,m}^{-1} \), \( \hat{B}_d = \hat{b}_d \hat{b}_d^T \), \( \hat{b}_d = \sum_{j \in N(i)} \tilde{w}_{ij} k_j^T K_{m,m}^{-1} \), \( \sigma_i^2 = k_{ii} - k_{i,m}^T K_{m,m}^{-1} k_i \), and \( \tilde{\sigma}_i = \sum_{j \in N(i)} \tilde{w}_{ij}^2 k_{jj} - k_j^T K_{m,m}^{-1} k_j \)

**Approximating** \( \mathbb{E}[\log p(c)] \):

Similar to the previous section, we approximate \( \mathbb{E}[\log p(c)] \) with the primary difference being that the mean \( \mu_i \) is computed from \( u \). Let \( \hat{\mu}_i = K_{i,m}^{-1} k_i \), then,

\[
\mathbb{E}[\log p(c)] \approx \sum_{i=1}^{|V|} \left[ (\alpha_i^0 - 1)(\log \hat{\mu}_i) - \beta_i \hat{\mu}_i \right].
\]

(18)

**Computing** \( \mathbb{E}[\log \int q(u)p(c|u)] \):

This expectation can be recognized as the entropy of a normal distribution given by:

\[
\tilde{H}(c) = \frac{|V|}{2}(1 + \log 2\pi) + \frac{1}{2} \log |\Sigma_c|
\]

(19)

where \( \tilde{\Sigma}_c \) is the covariance matrix obtained from the integration with \( q(u) \),

\[
\tilde{\Sigma}_c = K_n - K_{nm} K_{m,m}^{-1} K_{m,n} + K_{nm} K_{m,m}^{-1} S K_{m,m}^{-1} K_{m,n}.
\]
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 $\langle C_c \rangle$.

| Dataset | $|V|$ | $|E|$ | Attr. | Weights | $\langle k \rangle$ | $\langle C_c \rangle$ |
|---------|------|------|-------|--------|----------------|----------------|
| ER      | 100  | 1028 | –     | N      | 10.3          | 0.11           |
| BA      | 100  | 588  | –     | N      | 5.9           | 0.14           |
| CA      | 65   | 1139 | –     | N      | 35.0          | 0.13           |
| CE      | 239  | 3434 | –     | N      | 14.4          | 0.31           |
| FM      | 32   | 460  | 4     | Y      | 12.4          | 0.28           |
| FB      | 334  | 5704 | 4     | N      | 17.1          | 0.68           |
| PT      | 57   | 3249 | 9     | Y      | 48.4          | 1.00           |
| TX      | 150  | 22129| –     | Y      | 147.5         | 0.97           |

**Summarized Lower-Bound:** Putting all the terms together, along with preceding results leads us to a lower bound:

$$L_4(q) = -\frac{|D|}{2} \log \sigma_n^2 - \frac{1}{2\sigma_n^2} \sum_{d \in D} \Lambda_d + \sum_{i=1}^{|V|} \left[ (\alpha_i^0 - 1)(\log \tilde{\mu}_i) - \beta_i \tilde{\mu}_i \right] - \tilde{\mathbb{H}}(c) - \mathbb{D}_{KL}(q(\lambda) \| p(\lambda)) + \text{const}$$

Note that unlike $L_2$, we do not have to maintain separate distributions for each $c_i$. In the VBC-GP, the centralities are represented indirectly via $u$ and the learned mapping.

## 5 Experimental Results

In this section, we present results of experiments using VBC and VBC-GP to infer centrality in two synthetic and six real-world networks, i.e., the Erdős-Rényi (ER) random graph [17] and Babárai-Albert (BA) scale-free network [18], a cat brain connectome (CA) [19], the neural network of the *C. elegans* nematode (CE) [20, 21], a social network constructed from a Facebook ego user (FB) [22], Freeman’s EIES network of communication between 32 academics (FM) [23], the Singapore public-transportation people-flow network (PT) and a set of Taxi movement networks (TX). Basic properties of the networks are shown in Table 1. TX comprised daily taxi flows between 150 zones in Singapore for one week in 2010 and a one additional network of total flows for one month. Our models, datasets and experiment code are available and can be used to reproduce our results and figures. The lower-bounds were optimized using an interior point method with an initial point set using the eigenvector centrality of a noisy weight matrix.

---

1. See Supplementary Material and online at [http://www.haroldsoh.com](http://www.haroldsoh.com)
Figure 1: Median correlation scores (with upper and lower quartiles) for each of the tested networks for three noise levels and varying sample sizes. The Bayesian centralities began at a similar correlation score as the baseline (BL) and approached the noise-free centralities as more samples were provided. With 5 samples per node, VBC achieved high scores (\( \rho > 0.9 \)) across the networks, even under high noise conditions (\( \sigma^2_n = 1.0 \)).

5.1 Centrality with Noisy Edges

Our first set of experiments was designed to test if the VBC model was capable of estimating centralities given repeated noisy edge observations. For the TX network, we treated each daily network as a noisy observation and the month-long total flow network as the "true" network. For the remaining networks, we generated graphs having links with zero-mean Gaussian noise \( \sigma^2_o = 0.1, 0.5 \) and 1.0; note that this noise generation process is different from the specification in the model (\( \sigma^2_o \neq \sigma^2_n \)). For each noise level, we provided VBC with an increasing number of samples per node (\( N_s = 1, 3, 5, 7, 9 \)) and each of these runs was repeated 15 times. As a performance measure, we computed the average Pearson correlation to eigenvector centralities of the noise-free network.

Our results are summarized in Fig. 1. We see that with a single sample per node, the VBC infers centralities comparable to the baseline (BL), i.e., eigenvector centrality computed using a noisy weight matrix. With larger \( N_s \), the Bayesian centralities become increasingly correlated with the centralities of the true underlying network. Using all available samples, the scores were high (\( \rho \geq 0.95 \)) across all the networks. This improvement was observed even under the highest noise level (\( \sigma^2_n = 1.0 \)), which caused \( \approx 15\% \) of links to be removed. We further compared the VBC scores (when using broad priors) against the centralities obtained from averaged weight matrices and found the relative difference in correlation to be small; on average 0.00049 (sd. 0.00034) across the noise levels and networks. Note that unlike the averaging method, the VBC allows for the incorporation of prior knowledge and gives a posterior

\(^{2}\)We also computed the Spearman correlation and top-ten overlap score. All the scores were strongly-correlated and hence, only the Pearson score is shown here for conciseness.
Figure 2: The VBC-GP learned high-quality mappings—as indicated by the high correlation scores close to the full-GP model (dashed line)—with increasing scores as $N_s$ increased (FM and PT). Also, a decrease in the inter-quartile ranges suggests more robust performance. Scatter plots (on the right) show the predicted centralities and associated uncertainties.
distribution over the inferred centralities.

5.2 Mapping Attributes to Centralities

In this experiment, we test the VBC-GP’s ability to determine an attribute-centrality mapping. We worked with the FM, FB and PT networks, which possessed node attributes, and generated noisy datasets with moderate link noise ($\sigma_n^2 = 0.5$). As before, 15 runs were conducted for each sample size setting $N_s$. The VBC-GP was set to use a radial basis function (RBF) kernel with an initial lengthscale of $l = 15$ and $m = 10, 20$ and 30 inducing inputs for FM, PT and FB respectively. Following [24], the inducing input locations $Z$ were selected via k-means clustering.

To ascertain the informativeness of the node attributes and comparison purposes, we trained a full GP with hyperparameter optimization (no sparsity employed) on the true (noise-free) 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 Pearson correlation (averaged over 15 runs).

The scores obtained by full-GP (dashed line) and VBC-GP are shown in Fig. 2. The VBC-GP scores were close to that of the full-GP, i.e., marginally lower ($\approx 2\%$ at $N_s = 9$) for the social networks (FM and FB), and slightly higher for PT ($N_s \geq 5$). The VBC-GP mapping quality—indicated by the high correlation scores—improved as $N_s$ increased for FM and PT networks, but not for FB. We posit that the reason behind the lack of improvement on FB was that the VBC-GP (given 30 inducing inputs) had already reached its upper-bound performance at $N_s = 1$. That said, the decreasing inter-quartile ranges indicate more robust performance with higher $N_s$. Overall, the VBC-GP predicted centralities—using only 10 to 30 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. 2).

6 Discussion, Limitations and Future Work

The experimental results demonstrate the merits of the variational Bayesian approach, i.e., the ability to handle noisy links and assimilate additional sample observations to better estimate node centralities. In the case of VBC-GP, the results show that high-quality mappings could be learned in the presence of noise. The correlations compare favorably to the full-GP (trained using the noise-free centralities), with the benefits of (1) sparsity and (2) the centralities are computed along with the mapping in a single optimization process.

The proposed model is not without limitations. While the Gaussian likelihood worked well in our experiments, it may be inappropriate where the noise process results in different deviation distributions. In such cases, an alternative likelihood function can be substituted into (5). For example, a log-normal, $p(D|c, \lambda) = \prod \log N(\sum_{j \in N(i)} \hat{w}_{ij} c_j | \lambda c_i, \sigma_n i)$, may be more suitable (given non-negative weights), but the derivations require additional approximation steps.
due to logs in the expectations. Experimentation with different likelihoods under varying noise conditions would make interesting future work.

For simplicity, we assumed a single point-optimized noise term $\sigma_n^2$, which was initialized using $\sigma_n^2 = \mathcal{V}[\lambda' c' - \sum_{j \in N(i)} \tilde{w}_{ij} c'_j]$ where $\lambda'$ and $c'$ were computed using a noisy weight matrix. Although effective in our experiments, a way forward would be to reintroduce the noise variables $\sigma_n^2$ with suitable priors. As an alternative, a parametric solution may be found since the noise terms are correlated with centrality. For the VBC-GP, a heteroscedastic approach can be applied, e.g., using a second GP to model the noise.

As mentioned in Section 2, PageRank is a variant of eigenvector centrality with centralities weighted by the outgoing links, $c_{pg} = \frac{a}{\sum_j w_{ij}} c_{pg} + b$. Incorporating such a weighting into the likelihood model would result in a Bayesian form of PageRank centrality.

Finally, many real-world networks are larger than those used in this study (limited due to available computational resources). The sparse nature of the VBC-GP is promising in that it lends itself nicely to deriving centralities for very large networks; preliminary experiments using mini-batch optimization for larger BA networks (5000 nodes, ≈50,000 edges) achieved scores of 0.9030 (the baseline Full-GP score was 0.9039). Using the natural gradient [24] may further improve scalability.

7 Conclusions

In this paper, we contributed a Bayesian model of eigenvector centrality and derived two variational lower-bounds for finding approximate posterior centralities. The second bound incorporates a sparse GP that maps node-attributes to latent centralities, thus alleviating the need to maintain separate distributions for each node. Experiments show that the VBC and VBC-GP achieved strong results with higher correlation to the true centralities as more samples were provided. Taking a broader perspective, this work bridges metrics for complex networks with modern machine learning, and we hope that it will serve as a catalyst for the development of other network metric models.

References

[1] A. Özgür, T. Vu, G. Erkan, and D. R. Radev, “Identifying gene-disease associations using centrality on a literature mined gene-interaction network,” Bioinformatics, vol. 24, pp. i277–i285, 07 2008.

[2] S. P. Borgatti, “Identifying sets of key players in a social network,” Computational & Mathematical Organization Theory, vol. 12, no. 1, pp. 21–34, 2006.

[3] H. Soh, S. Lim, T. Zhang, X. Fu, G. K. K. Lee, T. G. G. Hung, P. Di, S. Prakasam, and L. Wong, “Weighted complex network analysis of travel
routes on the singapore public transportation system,” *Physica A*, vol. 389, no. 24, pp. 5852–5863, 2010.

[4] 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.

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

[6] 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.

[7] 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.

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

[9] 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.

[10] 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.

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

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

[13] O. Perron, “Zur theorie der matrices,” vol. 64, no. 2, pp. 248–263, 1907.

[14] G. Frobenius, “Uber Matrizen aus nicht negativen Elementen,” pp. 456–477, 1912.

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

[16] 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.

[17] P. Erdős and A. Rényi, “On the evolution of random graphs,” *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, vol. 5, pp. 17–61, 1960.
[18] A.-L. Barabási and R. Albert, “Emergence of scaling in random networks,” *Science*, vol. 286, no. 5439, pp. 509–512, 1999.

[19] M. A. de Reus and M. P. van den Heuvel, “Rich club organization and intermodule communication in the cat connectome,” *The Journal of Neuroscience*, vol. 33, no. 32, pp. 12929–12939, 2013.

[20] J. G. White, E. Southgate, J. N. Thomson, and S. Brenner, “The structure of the nervous system of the nematode caenorhabditis elegans,” *Phil. Trans. of the Royal Society of London B*, vol. 314, no. 1165, pp. 1–340, 1986.

[21] D. J. Watts and S. H. Strogatz, “Collective dynamics of ‘small-world’ networks,” *nature*, vol. 393, no. 6684, pp. 440–442, 1998.

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

[23] S. C. Freeman and L. C. Freeman, *The networkers network: a study of the impact of a new communications medium on sociometric structure*. School of Social Sciences University of Calif., 1979.

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