On Consistent Vertex Nomination Schemes

Vince Lyzinski   VLYZINSKI@UMASS.EDU
Department of Mathematics and Statistics
University of Massachusetts Amherst
Amherst, MA 01003, USA

Keith Levin   KLEVIN@UMICH.EDU
Department of Statistics
University of Michigan
Ann Arbor, MI 48109, USA

Carey E. Priebe   CEP@JHU.EDU
Department of Applied Mathematics and Statistics
Johns Hopkins University
Baltimore, MD 21218-2608, USA

Abstract

Given a vertex of interest in a network G_1, the vertex nomination problem seeks to find the corresponding vertex of interest (if it exists) in a second network G_2. A vertex nomination scheme produces a list of the vertices in G_2, ranked according to how likely they are judged to be the corresponding vertex of interest in G_2. The vertex nomination problem and related information retrieval tasks have attracted much attention in the machine learning literature, with numerous applications to social and biological networks. However, the current framework has often been confined to a comparatively small class of network models, and the concept of statistically consistent vertex nomination schemes has been only shallowly explored. In this paper, we extend the vertex nomination problem to a very general statistical model of graphs. Further, drawing inspiration from the long-established classification framework in the pattern recognition literature, we provide definitions for the key notions of Bayes optimality and consistency in our extended vertex nomination framework, including a derivation of the Bayes optimal vertex nomination scheme. In addition, we prove that no universally consistent vertex nomination schemes exist. Illustrative examples are provided throughout.

Keywords: Vertex nomination, graph inference, recommender systems

1. Introduction

Statistical inference on graphs is an important branch of modern statistics and machine learning. In recent years, there have been numerous papers in the literature developing graph analogues of statistical inference tasks such as hypothesis testing (Asta and Shalizi, 2015; Tang et al., 2017b), classification (Tang et al., 2013; Chen et al., 2016a), and clustering (Luxburg, 2007; Rohe et al., 2011; Sussman et al., 2012; Newman and Clauset, 2016). Moreover, growth in the size and complexity of network data sets have necessitated techniques for network-specific data mining tasks such as link prediction (Liben-Nowell
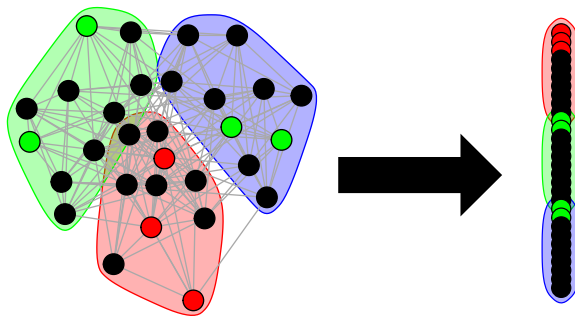
Figure 1: A visual representation of the classical Vertex Nomination framework: Given a community of interest in a network (here the red community) and some examples of vertices that are/are not part of the community of interest (colored red and green, respectively), rank the remaining vertices in the network into a nomination list, with those vertices from the community of interest concentrating at the top of the nomination list.

and Kleinberg, 2007; Lü and Zhou, 2011); entity resolution and network alignment (Conte et al., 2004; Lyzinski, 2018); and vertex nomination (Coppersmith and Priebe, 2012; Coppersmith, 2014; Suwan et al., 2015; Fishkind et al., 2015; Lyzinski et al., 2016). Akin to the development of classical statistics, algorithmic advancement has, in many ways, outpaced theoretical developments in these emerging graph-driven domains. This development has been necessitated by the dizzying pace of data generation, but there is nevertheless the need for a firm theoretical context in which to frame algorithmic progress. In this paper, drawing inspiration from the long-established classification framework in the pattern recognition literature (Devroye et al., 1997), we provide a rigorous theoretical framework for understanding statistical consistency in the vertex nomination inference task.

The vertex nomination (VN) task, which can be viewed as the graph analogue of the more classical recommender system task (Ricci et al., 2011), has traditionally been stated as follows: given a community of interest in a network and some examples of vertices that are or are not part of a community of interest, vertex nomination seeks to rank the remaining vertices in the network into a nomination list, with those vertices from the community of interest (ideally) concentrating at the top of the nomination list. See Figure 1 for a visual representation of this classical Vertex Nomination framework. In limited-resource settings, vertex nomination tools have proven to be effective in efficiently searching and querying large networks, with applications including detecting fraudsters in the Enron email network (Coppersmith and Priebe, 2012; Marchette et al., 2011; Suwan et al., 2015), uncovering web advertisements that have association with human trafficking (Fishkind et al., 2015), and identifying latent structure in connectome data (Fishkind et al., 2015; Yoder et al., 2018).

While related to the community detection problem (Newman, 2006; Luxburg, 2007; Bickel and Chen, 2009; Newman and Clauset, 2016), this traditional formulation of the VN problem is a semi-supervised inference task whose output is not an assignment of vertices to communities, but rather a ranked estimate of which vertices belong to a particular community of interest. That is, in contrast to community detection, the VN problem does

not aim to recover the community memberships of any vertices not in the community of interest. Clearly, any method that can recover the community memberships of all vertices in a graph can recover the interesting community, and hence any community detection algorithm can be repurposed for the VN problem just described with minor adaptation (e.g., by ranking vertices according to their probability of membership in the community of interest); see, for example, the spectral vertex nomination scheme of Fishkind et al. (2015). The specific performance of such an adaptation is highly dependent on the fidelity of the base clustering procedure, and the performance is often below that of the semi-supervised VN specific analogues (see Yoder et al., 2018).

The above formulation of the VN task assumes the presence of strong community structure among the vertices of interest in the graph. In practice, this is often a reasonable assumption, particularly if it is expected that interesting vertices will behave similarly to one another in the network. However, the particular features that mark a vertex as interesting are entirely task-dependent. To paraphrase the common proverb, interestingness is in the eye of the practitioner. Interesting vertices may be, for example, those with large network centrality (Jeong et al., 2001; Newman, 2005), those with a particular role in the network (Lusseau and Newman, 2004), or those corresponding to a given user across social networks (Patsolic et al., 2017). In these applications, interesting vertices need not correspond precisely to the community structure captured by a generative network model, and hence such cases are ill-described by the community-based VN problem described above. To accommodate this task-dependency and broader notion of interesting vertices, we consider the following generalization and extension of the previously-presented VN problem: Given a vertex of interest v^* in a graph $G_1 = (V_1, E_1)$, find the corresponding vertex of interest u^* (if it exists) in a second graph $G_2 = (V_2, E_2)$ by ranking the vertices of G_2 according to our confidence that they correspond to v^* in graph G_1; see Figure 2 for a visual representation of this VN framework. In this formulation, which is an (potentially) unsupervised inference task, what defines v^* as interesting is entirely model-dependent, and different network models can highlight different characteristics of interest in the graph. Potential application domains for this VN generalization abound, including identifying users of interest across social network platforms (see, for example, Patsolic et al., 2017), identifying structural signal across connectomes (see, for example, Sussman et al., 2018), and identifying topics of interest across graphical knowledge bases (see, for example, Sun and Priebe, 2013).

In Fishkind et al. (2015) and Lyzinski et al. (2016), the notion of a consistent vertex nomination scheme (i.e., an asymptotically optimal solution to the VN problem) was proposed for the original formulation of the VN problem, in which community membership entirely determines whether or not a given vertex is interesting. This definition of consistency was based on the *mean average precision* (MAP) of a nomination scheme operating on a graph model with explicit community structure encoded by the the Stochastic Block Model (SBM) of Holland et al. (1983). Under this restricted notion of consistency, Fishkind et al. (2015) derived the analogue of universal Bayes optimality in the VN setting, namely a scheme that achieves the optimal mean average precision for all parameterizations of the underlying SBM. While this derivation of the Bayes optimal scheme somewhat parallels the derivation of the Bayes optimal classifier in the classical pattern recognition literature, the SBM model assumption and MAP formulation greatly narrow the set of models and sets of interesting vertices we can consider. In this paper, we revamp and generalize the concept
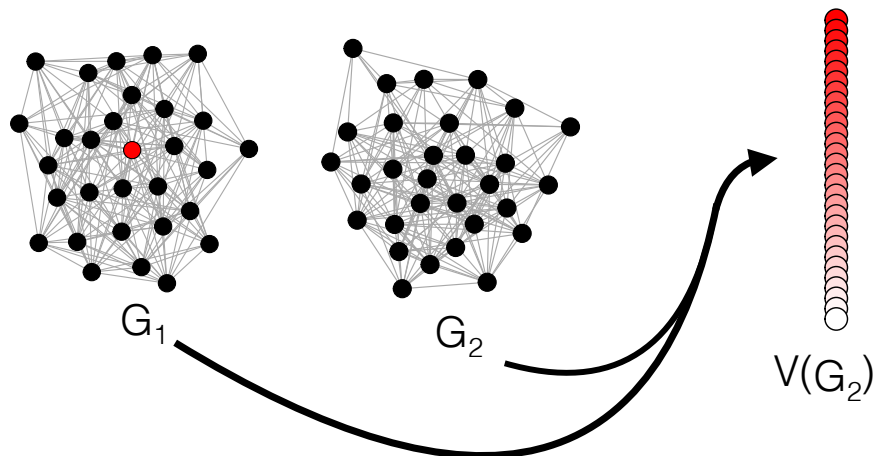
Figure 2: A visual representation of the generalized Vertex Nomination framework: Given a vertex of interest v^* (colored red) in a graph $G_1 = (V_1, E_1)$, find the corresponding vertex of interest u^* (if it exists) in a second graph $G_2 = (V_2, E_2)$, ranking the vertices of G_2 into a nomination list so that u^* ideally appears at the top of the nomination list.

of VN consistency—and of VN Bayes optimality—in the two-graph VN framework. This framework is quite general, and further allows us to highlight the similarities and differences between our new VN consistency formulation and its analogue in the classification literature defined in, for example, Devroye et al. (1997).

The paper is laid out as follows. In the remainder of this section, we provide brief overviews of information retrieval as it relates to vertex nomination (Section 1.1) and the Bayes optimal classifier in the classical setting (Section 1.2), and conclude the introduction by establishing notation for the remainder of the paper (Section 1.3). In Section 2, we define the VN problem framework that is the focus of this paper, and in Section 3 we derive the VN analogue of a Bayes optimal scheme. In Section 4, we define a new notion of VN consistency, and we prove that no universally consistent VN scheme exists, providing an interesting contrast to the standard classification setting. We conclude in Section 5 with a short summary comparing and contrasting VN with classical classification and a discussion of implications and future directions.

1.1. Connections to Information Retrieval

The vertex nomination task is, in some ways, similar to the task faced by recommender systems (Resnick and Varian, 1997; Ricci et al., 2011), in which the aim is to retrieve objects (e.g., documents or images) likely to be of interest to a user based on his or her previous behavior. For example, the celebrated PageRank algorithm (Brin and Page, 1998) recommends webpages based on random walks on the world wide web graph, in which websites are nodes and (directed) edges reflect hyperlinks between pages. The information retrieval (IR) literature includes many such graph-based approaches. We refer the reader to

Ricci et al. (2011) and Mihalcea and Radev (2011) for the state of the art circa 2010, and concentrate here on recapping more recent graph-based information retrieval techniques.

Many graph-based IR techniques rely on the assumption that similar objects (i.e., documents, webpages, etc.) will lie near one another in a suitably-constructed graph. Indeed, this intuition underlies many graph-based approaches throughout machine learning and related disciplines (see, for example, Belkin and Niyogi, 2003; Zhou et al., 2004). Techniques along these lines have been applied toward many tasks in natural language processing, typically inspired by PageRank (Rothe and Schütze, 2014). Along similar lines, Ma et al. (2012) applied a diffusion-based method (Coifman and Lafon, 2006) to the world wide web graph to yield an approach to ranking for query completion and recommendation. These information retrieval techniques can be naturally adapted to the vertex nomination problem by treating the vertex or vertices of interest as the object or objects to be retrieved.

The vertex nomination problem also bears similarities to the task of *learning to rank* (Duh, 2009; Liu, 2009; Li, 2011), in which the goal is to learn an ordering on a set of objects (e.g., documents, images, videos, etc.) according to (estimated) similarity or relevance to a given query object. In the learning to rank literature, graphs usually appear as training instances, with nodes corresponding to objects and edges encoding preferences or similarities among them elicited from users (e.g., an undirected weighted edge may join two documents judged to be similar). The work in Agarwal et al. (2006) is among the earliest to consider the problem of ranking objects in a network. The authors modified the PageRank algorithm to take preference information into account, rather than working solely with the hyperlink graph. In Agarwal (2010), the authors used a *data graph* encoding object similarities to obtain a regularizer similar to Belkin et al. (2006) on the empirical ranking error, with the target ranking encoded in a *preference graph*. More recent efforts along these lines have focused on the problem of incorporating network structure present between entities of different types, for example, between users and events in a social network (Luo et al., 2014; Pham et al., 2016). Here again, any learning to rank algorithm has a natural adaptation to the VN problem by using the first graph, in which some vertices are labeled, as training data to learn a ranking on the vertices of the second graph.

1.2. Bayes Error in Classical Pattern Recognition

In this section, we review the concepts of consistency and Bayes error from the statistical classification literature. We do not aim to give an exhaustive overview of the subject, but only to provide a rough outline as to the structures that we would like to replicate in the context of vertex nomination. For a more thorough treatment, we refer the interested reader to Devroye et al. (1997), whose presentation we follow below.

We begin by recalling the classical definition of Bayes error. Note that we will restrict our attention to the two-class problem to maximally bring forth the similarities (and differences) between statistical classification and VN, as in VN vertices are either of interest or not.

Definition 1 *Consider a set of potential observations $\mathcal{X}$ and a set of unknown class labels $\{0,1\}$ for objects in $\mathcal{X}$. A classifier is a function $h : \mathcal{X} \rightarrow \{0,1\}$, which aims to predict the class label of a given observation in $\mathcal{X}$. Given a distribution F supported on $\mathcal{X} \times \{0,1\}$, the error for the classifier h is given by $L(h) = \mathbb{P}(h(X) \neq Y)$ where $(X,Y) \sim F$.*

Any classifier that achieves the lowest possible error is said to be a *Bayes optimal* classifier. We write h^* for any such optimal classifier, which by definition satisfies $h^* \in \arg\min_{h:\mathcal{X}\to\{0,1\}} L(h)$. It is easily seen in this two-class framework that the Bayes optimal classifier is given by

$$h^*(x) = \begin{cases} 1 & \text{if } \mathbb{E}(Y|X=x) = \mathbb{P}(Y=1|X=x) > 1/2; \\ 0 & \text{else.} \end{cases} \tag{1}$$

Practically speaking, the Bayes optimal scheme chooses the label which maximizes the class-conditional probability of the observed data. The corresponding error, $L^* = L(h^*)$, is called the *Bayes error*. Of course, h^* depends on the distribution F of (X, Y), and, when appropriate, we will make this dependence explicit by writing L_F^*.

In practice, a classifier is often constructed based on training data $(X_1, Y_1), (X_2, Y_2),$ $\ldots, (X_n, Y_n)$, where the data (X_i, Y_i) are drawn i.i.d. according to F. This supervised classification framework is defined as follows.

Definition 2 *Consider a set of potential observations $\mathcal{X}$ and a set of unknown class labels $\{0, 1\}$ for objects in $\mathcal{X}$. A (supervised) classifier is a function*

$$h_n : \mathcal{X} \times \{\mathcal{X} \times \{0, 1\}\}^n \to \{0, 1\},$$

which aims to predict the class label of a given observation in $\mathcal{X}$ based on n training observations $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n) \in \mathcal{X} \times \{0, 1\}$. Given a distribution F supported on $\mathcal{X} \times \{0, 1\}$, the error for the classifier h_n is given by

$$L_F(h_n) = \mathbb{P}\big[h_n(X, (X_i, Y_i)_{i=1}^n) \neq Y \mid (X_i, Y_i)_{i=1}^n\big]$$

where $(X, Y), (X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n) \overset{i.i.d.}{\sim} F$.

Note that $L_F(h_n)$ is a random variable in which $\{(X_i, Y_i)\}_{i=1}^n$ are drawn i.i.d. from F, but then held fixed as we average over $(X, Y) \sim F$.

A sequence of classifiers $\mathbf{h} = (h_n)_{n=1}^\infty$ is called a *classification rule*. Informally, a good classification rule is one for which the probability of error becomes arbitrarily close to Bayes optimal as $n \to \infty$. The precise nature of what we mean by close is codified in the concept of statistical consistency.

Definition 3 *A classification rule $\mathbf{h} = (h_n)_{n=1}^\infty$ is* consistent *with respect to F if*

$$\mathbb{E}_F(L(h_n)) \to L_F^*.$$

The rule $\mathbf{h}$ is strongly consistent *if*

$$L_F(h_n) \overset{a.s.}{\to} L_F^*.$$

A rule that is (strongly) consistent for all distributions F on $\mathcal{X} \times \{0, 1\}$ is called (strongly) universally consistent.

Perhaps surprisingly, given that F can have arbitrary structure on $\mathcal{X} \times \{0,1\}$, universally consistent classification rules exist; see Stone (1977) for the first proof of this phenomenon.

In Fishkind et al. (2015), a notion of consistency for vertex nomination was presented, roughly analogous to Definition 3. In contrast to the classification task presented above, vertex nomination requires a ranking of the vertices, rather than merely the classification of a single vertex. As such, a vertex nomination scheme is evaluated in Fishkind et al. (2015) based on *average precision* (Manning et al., 2008), rather than simply a fraction of correctly-classified vertices. In Fishkind et al. (2015), VN consistency is defined in the context of stochastic block model (SBM) random graphs with respect to a provably optimal *canonical* nomination scheme. This canonical scheme plays an analogous role of Bayes optimal classifiers in this restricted model framework (see Section 3 below). The goal of this paper is to explore and further develop a broader notion of VN consistency that encompasses a more expressive class of models than the SBM.

1.3. Notation and Background

We conclude this section by establishing notation and reviewing a few of the more popular statistical network models that we will make use of as examples in the sequel.

1.3.1. NOTATION

For a set S, we let $|S|$ denote its cardinality and $\binom{S}{2}$ denote the set of all unordered pairs of distinct elements from S. Throughout, we will denote graphs via the ordered pair $G = (V, E)$, with vertices V and edges $E \subseteq \binom{V}{2}$. All graphs considered herein will be labeled, hollow (i.e., containing no self-edges), and undirected. We let $\mathcal{G}_n$ denote the set of all labeled, hollow, undirected graphs on n vertices. Given a graph G, we will let $V(G)$ denote the vertices of G and $E(G)$ denote its edges. We note that when G is random, this latter set is a random subset of $\binom{V}{2}$. For a set of vertices $S \subseteq V(G)$, we let $G[S]$ denote the subgraph of G induced by S, i.e., the graph $G' = (S, E)$ with $\{u, v\} \in E$ if and only if $\{u, v\} \in E(G)$. In a few places, we will require the notion of an *asymmetric* graph. A graph $G \in \mathcal{G}_n$ is *asymmetric* if it has no nontrivial automorphisms (Erdős and Rényi, 1963). For a positive integer $n \in \mathbb{Z}$, we will define $[n] = \{1, 2, \dots, n\}$, and $\mathcal{G}_n$ to be the be the set of labeled graphs on n vertices. Throughout this paper, we will often, in order to simplify notation, suppress dependence of parameters on n. Throughout, the reader should assume that, unless specified otherwise, all parameters depend on the number of vertices n.

1.3.2. STOCHASTIC BLOCK MODELS

The stochastic block model (SBM) is a widely studied model for edge-independent random graphs with latent community structure (Holland et al., 1983; Hoff et al., 2002; Karrer and Newman, 2011).

Definition 4 *We say that a random graph $G = (V, E) \in \mathcal{G}_n$ is an instantiation of a stochastic block model with parameters (K, B, b), written $G \sim \mathrm{SBM}(K, B, b)$, if*

    i. V is partitioned into K classes (called communities or blocks), $V = V_1 \cup V_2 \cup \cdots \cup V_K$.

    ii. The block membership vector $b \in [K]^{|V|}$ is such that for all $k \in [K]$, $b_v = k$ if and only if $v \in V_k$.

iii. *The symmetric matrix $B \in [0,1]^{K \times K}$ denotes the edge probabilities between and within blocks, with*

$$\mathbb{1}_{\{\,\{u,v\} \in E(G)\}} \stackrel{ind.}{\sim} Bernoulli(B_{b_u,b_v}).$$

We note that when $K = 1$, we recover the Erdős-Rényi random graph (Erdős and Rényi, 1959), in which the edges of G are present or absent independently with probability p. In this special case, we write $G \sim \mathrm{ER}(n, p)$. By a slight abuse of notation, for a symmetric matrix $P \in [0,1]^{n \times n}$, we will write $G \sim \mathrm{ER}(P)$ if, identifying the vertices of G with $[n]$, we have $\{i, j\} \in E(G)$ with probability $P_{i,j}$ independent of the other edges. With no restrictions on P, $\mathrm{ER}(P)$ random graphs can be viewed as n-block SBMs and are the most general edge-independent random graph model.

The latent community structure inherent to SBMs makes them a natural model for use in the traditional vertex nomination framework. Recall the traditional VN task: given a community of interest in a network and some examples of vertices that are or are not part of the community of interest, vertex nomination seeks to rank the remaining vertices in the network into a nomination list, with those vertices from the community of interest (ideally) concentrating at the top of the nomination list. As a result, previous work on VN consistency (Fishkind et al., 2015) has been posed within the SBM framework, with the optimal scheme only obtaining its optimality for SBMs. We note that we consider herein the SBM setting where communities are disjoint and each vertex can only belong to a single community. However, the results contained herein translate immediately to the mixed membership SBM setting (Airoldi et al., 2008); details are omitted for brevity.

1.3.3. Random dot product graphs

In stochastic block models, the block assignment vector can be viewed as a latent feature vector for the vertices in the network, with these features (i.e., block memberships) defining the connectivity structure in the network. The random dot product graph (RDPG) model (Young and Scheinerman, 2007) allows for more nuanced vertex features to be incorporated into the model and has been used as the setting for a VN formulation similar to the one proposed here; see Patsolic et al. (2017) for details.

Definition 5 *We say that a random graph $G = (V, E) \in \mathcal{G}_n$ is an instantiation of a d-dimensional random dot product graph with parameters X, written $G \sim \mathrm{RDPG}(X)$, if*

i. *The matrix $X \in \mathbb{R}^{n \times d}$ is such that $0 \leq (XX^T)_{i,j} \leq 1$ for all $i, j \in [n]$. The rows of X provide the latent features for the vertices in V.*

ii. *The edges of G are present or absent independently, with $\{i, j\} \in E(G)$ with probability $(XX^T)_{i,j}$. Written succinctly, $G \sim \mathrm{ER}(XX^T)$.*

We can view the RDPG model as a example of the more general *latent position* random graph model (Hoff et al., 2002), in which edge probabilities are determined by hidden vertex-level geometry.

Estimating the latent position structure in RDPGs is particularly amenable to spectral methods, and this model has played a prominent role in recent theoretical developments of spectral graph methods (see, for example, Rohe et al., 2011; Sussman et al., 2012; Tang

et al., 2017a). Note that the RDPG can be extended to a broader class of models, in which edge probabilities are given by evaluating a positive definite link function at vertices' latent positions as in, for example, Tang et al. (2013). While incorporating this more general family of latent position graphs into the present VN framework would be straightforward, we restrict our focus to the RDPG model of Definition 5 for ease of exposition.

1.3.4. CORRELATION ACROSS NETWORKS

The vertex nomination problem we consider in this paper presupposes the existence of a vertex of interest in a network G_1 and, ideally, a corresponding vertex of interest in a second network G_2. Often, such correspondences across networks are encoded into random graph models via edge-wise graph correlation (see, for example, Fishkind et al., 2019). Arguably the simplest such structure is seen in the ρ-correlated Erdős-Rényi model (see, for example, Lyzinski et al., 2015).

Definition 6 *We say that bivariate random graphs $(G_1, G_2) \in \mathcal{G}_n \times \mathcal{G}_n$ are an instantiation of a ρ-correlated $\mathrm{ER}(P)$ model, written $(G_1, G_2) \sim \rho\text{-}\mathrm{ER}(P)$, if*

  i. *Marginally, $G_1 \sim \mathrm{ER}(P)$ and $G_2 \sim \mathrm{ER}(P)$.*

  ii. *Edges are independent across G_1 and G_2 except that the indicators of the events $\{u, v\} \in E(G_1)$ and $\{u, v\} \in E(G_2)$ are jointly distributed as a pair of Bernoulli random variables with success probability $P_{u,v}$ and correlation ρ. If the correlation is allowed to vary across edges, so that these two events have correlation $\rho_{u,v}$, then collecting these correlations in a symmetric matrix $R = [\rho_{i,j}]_{i,j=1}^n$, we write $(G_1, G_2) \sim R\text{-}\mathrm{ER}(P)$ (see Lyzinski and Sussman, 2018).*

Ranging the values in R from 0 to 1 allows for the consideration of graphs that range from independent ($R = 0$) to isomorphic ($R = 1$). Intermediate values of R allow for the encoding of a correspondence across networks between these two extremes. We will also consider $R < 0$, in which case edges across networks are anti-correlated. This is particularly useful for modeling situations in which corresponding vertices stochastically behave differently across networks.

2. Vertex Nomination

Loosely stated, the vertex nomination problem we consider in this paper can be summarized as follows: Given a vertex of interest v^* in a graph $G_1 = (V_1, E_1)$, find the corresponding vertex of interest u^* (if it exists) in a second graph $G_2 = (V_2, E_2)$ by ranking the vertices of G_2 according to our confidence that they correspond to v^* in graph G_1. To formally define this version of vertex nomination, we will need to consider distributions on graphs with partially-overlapping node sets that have a built-in notion of vertex correspondence across graphs. To this end, we will consider distributions on $\mathcal{G}_n \times \mathcal{G}_m$, where $\mathcal{G}_n$ is the set of labeled graphs on n vertices, with vertex labels given by $\{v_1, v_2, \ldots, v_n\}$, and $\mathcal{G}_m$ is the set of labeled graphs on m vertices, with vertex labels given by $\{u_1, u_2, \ldots, u_m\}$. Note that for $i \in [n] \cap [m]$, v_i and u_i are merely vertex labels, and it is not necessarily the case that $v_i = u_i$. We follow this labeling convention in order to emphasize the reality that the vertex sets of G_1 and G_2 may only partially overlap.

Definition 7 (Nominatable Distributions) *For a given $n, m \in \mathbb{Z} > 0$, the set of* Nominatable Distributions *of order (n, m), denoted $\mathcal{N}_{n,m}$, is the collection of all families of distributions of the following form*

$$\mathbf{F}_{\Theta}^{(n,m)} = \{F_{c,\theta}^{(n,m)} \quad s.t. \ 0 \leq c \leq \min(n, m) \in \mathbb{Z}, \theta \in \Theta \subset \mathbb{R}^{d(n,m)}\}$$

where $F_{c,\theta}^{(n,m)}$ is a distribution on $\mathcal{G}_n \times \mathcal{G}_m$ parameterized by $\theta \in \Theta$ satisfying:

i. *The vertex sets $V_1 = \{v_1, v_2, \ldots, v_n\}$ and $V_2 = \{u_1, u_2, \ldots, u_m\}$ satisfy $v_i = u_i$ for $0 < i \leq c$. We refer to $C = \{v_1, v_2, \ldots, v_c\} = \{u_1, u_2, \ldots, u_c\}$ as the* core vertices. *These are the vertices that are shared across the two graphs and imbue the model with a natural notion of corresponding vertices.*

ii. *Vertices in $J_1 = V_1 \setminus C$ and $J_2 = V_2 \setminus C$, satisfy $J_1 \cap J_2 = \emptyset$. We refer to J_1 and J_2 as* junk *vertices. These are the vertices in each graph that have no corresponding vertex in the other graph.*

iv. *The induced subgraphs $G_1[J_1]$ and $G_2[J_2]$ are conditionally independent given θ.*

A few examples will serve to illustrate this definition. We will return to the three example settings below several times throughout the rest of the paper in order to highlight and illustrate phenomena of interest.

Example 1 (R-ER(P)) Let $(G_1, G_2) \sim R$-ER(P) with $P, R \in \mathbb{R}^{n \times n}$ and $R > 0$ entrywise, so that G_1 and G_2 have correlated edges as described in Section 1.3.4. In this example, the model parameter is $\theta = (P, R)$, and the vertex sets of the two graphs can be thought of as fully overlapping, i.e., $V_1 = V_2 = C = [n]$ and $J_1 = J_2 = \emptyset$, since the correlation structure conveyed in the entries of R encodes an explicit correspondence between the edges of G_1 and the edges of G_2 (and hence also a correspondence between V_1 and V_2). Note that if we consider $C = [k]$ with $k < n$, then we would require (after suitably ordering the vertices) $R_{u,v} = 0$ for $u, v > k$. This highlights the way in which θ (and hence the distribution $F_{c,\theta}^{(n,m)}$) can vary with c, and vice-versa.

Example 2 (RDPG) Let $m > n$ and suppose that $Y \in \mathbb{R}^{m \times d}$ has distinct rows and satisfies $(YY^T)_{i,j} \in [0, 1]$ for all $i, j \in [m]$. Let $X \in \mathbb{R}^{n \times d}$ be a submatrix of Y, and consider $G_1 \sim \text{RDPG}(X)$ and $G_2 \sim \text{RDPG}(Y)$, where G_1 and G_2 are conditionally independent given Y. In this example, we can consider $\theta = Y$, $V_1 = [n] = C$, $J_1 = \emptyset$, $V_2 = [n] \cup J_2$, and $J_2 = \{u_{n+1}, u_{n+2}, \ldots, u_m\}$. Note that as G_1 and G_2 are conditionally independent given θ, we could also consider $0 < c < n$ here as well. This illustrates that θ need not necessarily vary with c, and hence $F_{c,\theta}^{(n,m)}$ need not vary with c, either.

Example 3 (Independent Erdős-Rényi graphs) Let (G_1, G_2) be independent ER(n, p) random graphs. In this example, we can consider any $c \in [n]$. Note that if $c = 0$ here, then there is no corresponding vertex of interest in G_2, and this example serves as a natural boundary case between models in which nomination is possible and those in which it is not. As we will see below in Theorem 23, $c > 0$ may still yield chance performance for any nomination scheme, and the existence of a vertex correspondence does not necessarily imply any performance guarantees.

Remark 8 In addition to the edge-independent and conditionally edge-independent network models considered above, the class of nominatable distributions contains a host of other popular random graph models, including the Exponential Random Graph Model (Frank and Strauss, 1986; Snijders et al., 2006; Robins et al., 2007), the preferential attachment model (Albert and Barabási, 2002), and the Watts-Strogatz small world model (Watts and Strogatz, 1998), among others. Indeed, if we consider the case where $c = n = m$, then any parametric distribution on $\mathcal{G}_n \times \mathcal{G}_n$ is a nominatable distribution.

Remark 9 The core vertices C in a nominatable distribution correspond to the vertices that can be sensibly identified across graphs. Note that this set does not require any further structure, aside from the conditional independence of $G_1[J_1]$ and $G_2[J_2]$ given the parameter θ. Thus, we are largely free to specify any notion of correspondence we please. Depending on the application, this correspondence may be that of vertices playing similar structural roles, belonging to the same community, or some more complicated application-specific notion of correspondence. That is, the notion of cross-graph correspondence, and hence the notion of vertex similarity, is largely left to the practitioner to specify when she or he specifies an appropriate random graph model.

Given a pair of graphs $(G_1, G_2) \sim F_{c,\theta}^{(n,m)} \in \mathcal{N}_{n,m}$, if the vertices in C are known across graphs then identifying the corresponding vertex to $v^* \in C$ is immediate from the vertex labels. In practice, this information is unknown and the correspondences across graphs are only partially observed or even unobserved entirely. To model this added uncertainty, we consider passing the vertex labels of G_2 through an obfuscating function.

Definition 10 *Let $(G_1, G_2) \sim F_{c,\theta}^{(n,m)} \in \mathcal{N}_{n,m}$. An obfuscating function $\mathfrak{o} : V_2 \to W$ is a bijection from V_2 to W with $W \cap V_i = \emptyset$ for $i = 1, 2$. We call a set W satisfying $W \cap V_i = \emptyset$ for $i = 1, 2$ an obfuscating set, and for a given obfuscating set W, we let $\mathfrak{D}_W$ be the set of all obfuscating functions $\mathfrak{o} : V_2 \to W$.*

Here, $\mathfrak{o}$ models the practical reality that the correspondence of labels across graph is unknown a priori. Note that to ease notation, we shall write $\mathfrak{o}(G_2)$ (resp., $\mathfrak{o}(g_2)$ and $\mathfrak{o}(\mathcal{G}_m)$) to denote the graph G_2 (respectively, g_2 and $\mathcal{G}_m$) whose labels have been obfuscated via $\mathfrak{o}$.

Before defining a VN scheme, we must make one additional definition: for a graph $g \in \mathcal{G}_m$ and $u \in V(g)$, define

$$\mathcal{I}(u; g) = \{w \in V(g) \text{ s.t. } \exists \text{ an automorphism } \sigma \text{ of } g, \text{ s.t. } \sigma(u) = w\}.$$

Note that by taking σ to be the identity, we have $u \in \mathcal{I}(u; g)$. The vertices in $\mathcal{I}(u; g)$ are those that are, in a sense, topologically equivalent to the vertex u in g, and hence, in the absence of labels, indistinguishable from one another. As such, any sensibly-defined vertex nomination scheme should view all vertices in $\mathcal{I}(u; g)$ as being equally good matches to a vertex of interest v^*. Thus, a well-defined VN scheme should be "label-independent" in the following sense: The set of ranks of each set of equivalent vertices (i.e., each $\mathcal{I}(u; g_2)$) needs to be invariant to the particular choice of obfuscating function; see Figure 3 for an illustration of this consistency criterion. Formally, we have the following.
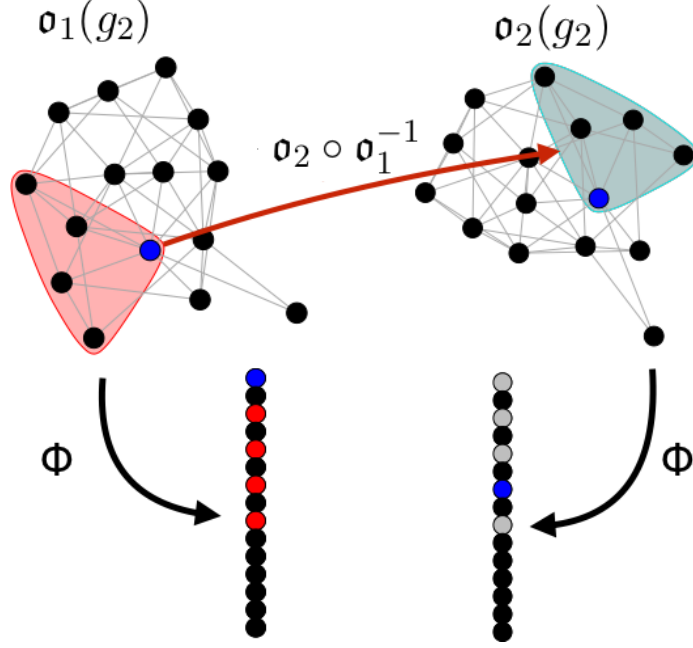
11

Figure 3: An illustration of the "label-independence" property of VN schemes. If the blue vertex in $\mathfrak{o}_1(g_2)$ (resp., $\mathfrak{o}_2(g_2)$) is $\mathfrak{o}_1(u)$ (resp., $\mathfrak{o}_2(u)$) for $u \in V_2$, then we require the ranks of $\mathcal{I}(\mathfrak{o}_1(u); \mathfrak{o}_1(g_2))$ (outlined in red in the network $\mathfrak{o}_1(g_2)$ and colored red/blue in the ordering provided by Φ) to be equal to the ranks of $\mathcal{I}(\mathfrak{o}_2(u); \mathfrak{o}_2(g_2))$ (outlined in grey in the network $\mathfrak{o}_2(g_2)$ and colored grey/blue in the ordering provided by Φ). Indeed, the set of ranks of $\mathfrak{o}(\mathcal{I}(u; g_2))$ via Φ is independent of the choice of obfuscation function $\mathfrak{o}$.

Definition 11 (Vertex Nomination (VN) Scheme) *For a set A, let $\mathcal{T}_A$ denote the set of all total orderings of the elements of A. For $n, m > 0$ fixed, obfuscating set W, and obfuscating function $\mathfrak{o} \in \mathfrak{D}_W$, a vertex nomination scheme is a function $\Phi : \mathcal{G}_n \times \mathfrak{o}(\mathcal{G}_m) \times V_1 \to \mathcal{T}_W$ satisfying the following consistency property: If for each $u \in V_2$, we define $rank_{\Phi(g_1, \mathfrak{o}(g_2), v^*)}(\mathfrak{o}(u))$ to be the position of $\mathfrak{o}(u)$ in the total ordering provided by $\Phi(g_1, \mathfrak{o}(g_2), v^*)$, and we define $\mathfrak{r}_\Phi : \mathcal{G}_n \times \mathcal{G}_m \times \mathfrak{D}_W \times V_1 \times 2^{V_2} \mapsto 2^{[m]}$ via*

$$\mathfrak{r}_\Phi(g_1, g_2, \mathfrak{o}, v^*, S) = \{rank_{\Phi(g_1, \mathfrak{o}(g_2), v^*)}(\mathfrak{o}(u)) \ s.t. \ u \in S\},$$

then we require that for any $g_1 \in \mathcal{G}_n$, $g_2 \in \mathcal{G}_m$, $v^* \in V_1$, obfuscating functions $\mathfrak{o}_1, \mathfrak{o}_2 \in \mathfrak{D}_W$ and any $u \in V(g_2)$,

$$\mathfrak{r}_\Phi\big(g_1, g_2, \mathfrak{o}_1, v^*, \mathcal{I}(u; g_2)\big) = \mathfrak{r}_\Phi\big(g_1, g_2, \mathfrak{o}_2, v^*, \mathcal{I}(u; g_2)\big) \tag{2}$$

$$\Leftrightarrow \mathfrak{o}_2 \circ \mathfrak{o}_1^{-1}\Big[\mathcal{I}\Big(\Phi(g_1, \mathfrak{o}_1(g_2), v^*)[k]); \mathfrak{o}_1(g_2)\Big)\Big] = \mathcal{I}\Big(\Phi(g_1, \mathfrak{o}_2(g_2), v^*)[k]; \mathfrak{o}_2(g_2)\Big), \ \text{for all } k \in [m],$$

where $\Phi(g_1, \mathfrak{o}(g_2), v^*)[k]$ denotes the k-th element (i.e., the rank-k vertex) in the ordering $\Phi(g_1, \mathfrak{o}(g_2), v^*)$. We let $\mathcal{V}_{n,m}$ denote the set of all such VN schemes.

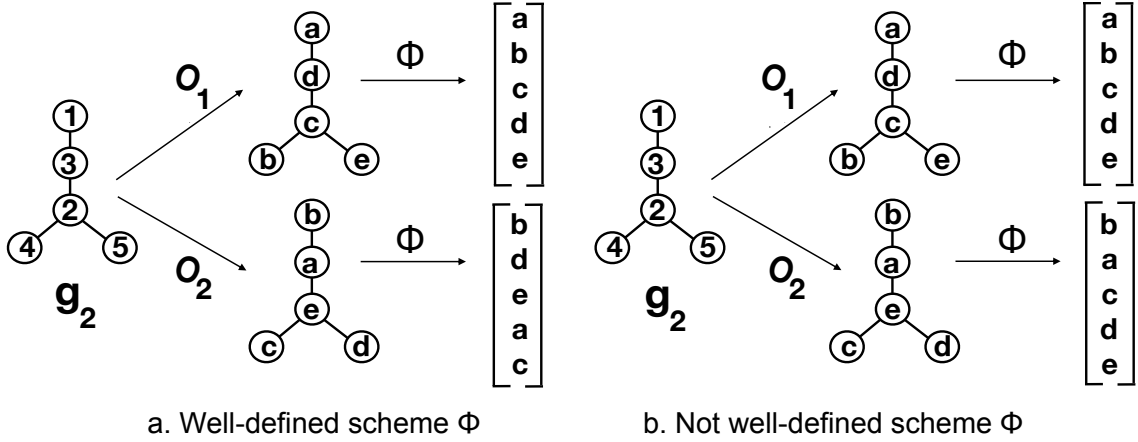a. Well-defined scheme Φ   b. Not well-defined scheme Φ

Figure 4: An example of the consistency criterion, Equation (2), in action. The left panel (a) shows a well-defined nomination scheme while the right panel (b) shows an ill-defined scheme. The key in this example is that any scheme satisfying Equation (2) must have $\mathfrak{r}_\Phi(g_1, g_2, \mathfrak{o}_1, v^*, \{1\}) = \mathfrak{r}_\Phi(g_1, g_2, \mathfrak{o}_2, v^*, \{1\})$; $\mathfrak{r}_\Phi(g_1, g_2, \mathfrak{o}_1, v^*, \{3\}) = \mathfrak{r}_\Phi(g_1, g_2, \mathfrak{o}_2, v^*, \{3\})$; $\mathfrak{r}_\Phi(g_1, g_2, \mathfrak{o}_1, v^*, \{2\}) = \mathfrak{r}_\Phi(g_1, g_2, \mathfrak{o}_2, v^*, \{2\})$; and $\mathfrak{r}_\Phi(g_1, g_2, \mathfrak{o}_1, v^*, \{4, 5\}) = \mathfrak{r}_\Phi(g_1, g_2, \mathfrak{o}_2, v^*, \{4, 5\})$.

Given $(G_1, G_2) \sim F_{c,\theta}^{(n,m)} \in \mathcal{N}_{n,m}$ realized as $G_1 = g_1$ and $G_2 = g_2$ with $v^* \in V_1$ the vertex of interest in G_1, a VN scheme $\Phi(\cdot, \cdot, \cdot)$ produces a ranked list $\Phi(g_1, \mathfrak{o}(g_2), v^*)$ of the vertices of $\mathfrak{o}(g_2)$ (i.e., the set W), ordered according to how likely each vertex in $V(\mathfrak{o}(g_2))$ is judged to correspond to v^*, with optimal performance corresponding to

$$\Phi(g_1, \mathfrak{o}(g_2), v^*)[1] = \begin{cases} \mathfrak{o}(v^*) & \text{if } v^* \in C \\ \text{arbitrary } v \in W & \text{if } v^* \notin C. \end{cases}$$

Less formally, one can think of a VN scheme as ranking the vertices of G_2 according to how well they resemble the vertex of interest v^* under some task-dependent measure.

Remark 12 Note that if $u \in V_2$ is such that $\mathcal{I}(u; g_2) = \{u\}$ (i.e., u is topologically distinct within g_2), then Equation (2) implies that

$$\text{rank}_{\Phi(g_1, \mathfrak{o}_1(g_2), v^*)}\big(\mathfrak{o}_1(u)\big) = \text{rank}_{\Phi(g_1, \mathfrak{o}_2(g_2), v^*)}\big(\mathfrak{o}_2(u)\big)$$

for any $\mathfrak{o}_1$, $\mathfrak{o}_2$ in $\mathfrak{O}_W$. If $\mathcal{I}(u; g_2)$ contains vertices in addition to u, then Equation (2) implies that the set of vertices topologically equivalent to u (namely, those in $\mathcal{I}(u; g_2)$) must achieve the same ranks via Φ under any two obfuscating functions; see Figure 4 for a simple example of this consistency criterion in action.

Remark 13 (Relation to Fishkind et al. (2015); Lyzinski et al. (2016)) Recall the one-graph vertex nomination task considered in earlier works (Coppersmith, 2014; Fishkind et al., 2015; Lyzinski et al., 2016) and described in Section 1, in which vertices are

13

considered interesting precisely when they belong to one of K communities in a stochastic block model. While the two-graph VN formulation we consider in the present work (modulo symmetries) involves a single vertex of interest across graphs, the framework is easily extended to the setting where one may have multiple vertices of interest (and not of interest), and in particular it can encode instances of the one-graph version VN problem. To see this, consider an instance of the single-graph VN problem on graph $G = (V, E)$ where V is partitioned into K communities as $V = V_1 \cup V_2 \cup \cdots \cup V_K$ and each of the communities is comprised of labeled (i.e., seed vertices, whose community memberships are observed) and unlabeled (i.e., nonseed, whose community memberships are unobserved) vertices, $V_k = S_k \cup U_k$, where $S_k \subseteq V_k$ is the set of seeds from the k-th block and $U_k \subseteq V_k$ is the set of nonseed vertices. We can encode this one-graph VN instance as an instance of the two-graph problem by encoding additional information in the graph G_1. Construct a vertex set $V' = V \cup \{\ell_1, \ell_2, \ldots, \ell_K\}$. The K new vertices $\{\ell_k\}_{k=1}^K$ will encode the label information present in the graph G. Let $E' = E \cup L$, where $L = \{\{\ell_k, s\} : s \in S_k, k \in [K]\}$, so that edges connect from seed vertices in $S = S_1 \cup S_2 \cup \cdots \cup S_K$ to their corresponding label vertices. Take $G_1 = (V', E')$, and let the interesting vertices (and possible uninteresting vertices) be given by the elements of $S \subseteq V'$. The second graph G_2 is then the subgraph of G induced by the unlabeled vertices $U \subseteq V$ passed through an appropriate obfuscating function. This pair (G_1, G_2), with any $s \in S_1$ chosen to be the interesting vertex, encodes the label information present in the one-graph VN problem as well as the graph structure of G, as required.

3. Bayes Error and Bayes Optimality in Vertex Nomination

Viewing a VN scheme as an information retrieval system suggests that a scheme that puts $\mathfrak{o}(v^*)$ close to the top of the nomination list is potentially of great practical value, even if it fails to obtain perfect performance. Motivated by this, we adapt the recall-at-k metric from classical information retrieval as a measure of performance. To wit, we define the level-k loss function and error for VN as follows.

Definition 14 (VN loss function, level-k error) *Let $\Phi \in \mathcal{V}_{n,m}$ be a vertex nomination scheme and $\mathfrak{o}$ an obfuscating function. For (g_1, g_2) realized from $(G_1, G_2) \sim F_{c,\theta}^{(n,m)} \in \mathcal{N}_{n,m}$ with vertex of interest $v^* \in C$, and for $k \in [m-1]$, we define the* level-k *nomination loss via*

$$
\begin{aligned}
\ell_k(\Phi, g_1, g_2, \mathfrak{o}, v^*) &= \mathbb{1}\{rank_{\Phi(g_1, \mathfrak{o}(g_2), v^*)}(\mathfrak{o}(v^*)) \geq k+1\} \\
&= 1 - \mathbb{1}\{rank_{\Phi(g_1, \mathfrak{o}(g_2), v^*)}(\mathfrak{o}(v^*)) \leq k\}.
\end{aligned}
\tag{3}
$$

The level-k *error of Φ at v^* is then defined to be*

$$
\begin{aligned}
L_k(\Phi, v^*) &= \mathbb{E}_{(G_1, G_2) \sim F_{c,\theta}^{(n,m)}} \left[\ell_k(\Phi, G_1, G_2, \mathfrak{o}, v^*)\right] \\
&= \mathbb{P}_{(G_1, G_2) \sim F_{c,\theta}^{(n,m)}} \left[rank_{\Phi(G_1, \mathfrak{o}(G_2), v^*)}(\mathfrak{o}(v^*)) \geq k+1\right].
\end{aligned}
\tag{4}
$$

From the definition of the level-k error in Eq. (4), it is immediate that

$$L_1(\Phi, v^*) = 1 - \mathbb{P}_{(G_1,G_2)\sim F_{c,\theta}^{(n,m)}} \left[\mathrm{rank}_{\Phi(G_1,\mathfrak{o}(G_2),v^*)}(\mathfrak{o}(v^*)) = 1\right]$$

$$\geq L_2(\Phi, v^*) = 1 - \mathbb{P}_{(G_1,G_2)\sim F_{c,\theta}^{(n,m)}} \left[\mathrm{rank}_{\Phi(G_1,\mathfrak{o}(G_2),v^*)}(\mathfrak{o}(v^*)) \in \{1,2\}\right]$$

$$\geq L_3(\Phi, v^*) = 1 - \mathbb{P}_{(G_1,G_2)\sim F_{c,\theta}^{(n,m)}} \left[\mathrm{rank}_{\Phi(G_1,\mathfrak{o}(G_2),v^*)}(\mathfrak{o}(v^*)) \in \{1,2,3\}\right] \qquad (5)$$

$$\vdots$$

$$\geq L_{m-1}(\Phi, v^*) = 1 - \mathbb{P}_{(G_1,G_2)\sim F_{c,\theta}^{(n,m)}} \left[\mathrm{rank}_{\Phi(G_1,\mathfrak{o}(G_2),v^*)}(\mathfrak{o}(v^*)) \in [m-1]\right],$$

The level-1 loss function is analogous to the classical 0/1 loss function in classification, as $L_1(\Phi, v^*)$ is simply the probability that Φ fails to "classify" $\mathfrak{o}(v^*)$ as the vertex corresponding to v^* in $\mathfrak{o}(G_2)$ (i.e., fails to rank it first). Considering $1 < k \ll m$ enables us to model the practical loss associated with using a VN scheme to search for $\mathfrak{o}(v^*)$ in $\mathfrak{o}(V_2)$ given limited resources.

Remark 15 Unlike in the classification setting described in Section 1.2, where $L_F(h_n)$ is a random variable indexed by n, the nomination errors defined in Definition 14 are sequences indexed by m and n and are not random. In the classical setting, $L_F(h_n)$ denotes the error rate of a classifier that classifies a single observation X based on n training instances $\{(X_i, Y_i)\}_{i=1}^n$. In the case of VN, the notion of labeled training instances is, at best, more hazy. Indeed, in the present setting, the training data and test data are inseparable. The graphs (or, more specifically, their edges) *are* the training data, and in the present work, the graph orders n, m are better thought of as measuring problem dimension rather than training set size.

Analogous to the classification literature, we are now able to define the concept of Bayes optimality in the VN framework.

Definition 16 (Bayes error of a VN scheme) *Let $(G_1, G_2) \sim F_{c,\theta}^{(n,m)}$ with vertex of interest $v^* \in C$ (where we recall that C is the set of core vertices; see Definition 7), and let $\mathfrak{o} \in \mathfrak{D}_W$ be an obfuscating function. For $k \in [m-1]$, we define the* level-k Bayes optimal *VN scheme to be any element $\Psi \in \arg\min_{\Phi \in \mathcal{V}_{n,m}} L_k(\Phi, v^*)$, and define the* level-$k$ Bayes error *to be $L_k^*(v^*) = L_k(\Psi, v^*)$ for level-k Bayes optimal Ψ.*

Now that we have a notion of Bayes error for VN, it is natural to ask whether an optimal VN scheme exists analogous to the Bayes optimal classifier of Equation (1). Toward this end, let (g_1, g_2) be realized from $(G_1, G_2) \sim F_{c,\theta}^{(n,m)} \in \mathcal{N}_{n,m}$, and consider a vertex of interest $v^* \in C$ and obfuscating function $\mathfrak{o} : V_2 \to W$. In order to avoid the technical complexities associated with graph automorphisms, in what follows we will assume that $F_{c,\theta}^{(n,m)} \in \mathcal{N}_{n,m}$ is supported on $\mathcal{G}_n^a \times \mathcal{G}_m^a$, where $\mathcal{G}_n^a$ (resp., $\mathcal{G}_m^a$) is the set of asymmetric graphs in $\mathcal{G}_n$ (resp., $\mathcal{G}_m$). For analogous results in networks with symmetries, see Remark 18.

Letting $\simeq$ denote graph isomorphism, for $(g_1, g_2) \in \mathcal{G}_n \times \mathcal{G}_m$ define the set

$$\begin{aligned} (g_1, [\mathfrak{o}(g_2)]) &= \left\{ (g_1, \tilde{g}_2) \text{ s.t. } \mathfrak{o}(\tilde{g}_2) \simeq \mathfrak{o}(g_2) \right\} \\ &= \left\{ (g_1, \tilde{g}_2) \text{ s.t. } \tilde{g}_2 \simeq g_2 \right\}. \end{aligned} \qquad (6)$$

In order to define the Bayes optimal scheme, we will also need the following restriction of $(g_1, [\mathfrak{o}(g_2)])$: for each $w \in W$, we define

$$(g_1, [\mathfrak{o}(g_2)])_{w=\mathfrak{o}(v^*)} = \left\{ (g_1, \tilde{g}_2) \text{ s.t. } \exists \text{ isomorphism } \sigma \text{ s.t. } \mathfrak{o}(\tilde{g}_2) = \sigma(\mathfrak{o}(g_2)), \sigma(w) = \mathfrak{o}(v^*) \right\}. \tag{7}$$

We are now ready to define a Bayes optimal VN scheme.

For ease of notation, in the sequel we will write $\mathbb{P}_{F_{c,\theta}^{(n,m)}}$ or even simply $\mathbb{P}$ in place of $\mathbb{P}_{(G_1,G_2) \sim F_{c,\theta}^{(n,m)}}$ where there is no risk of ambiguity. Let

$$\mathbf{g} = \left\{ \left(g_1^{(i)}, g_2^{(i)} \right) \right\}_{i=1}^{h} \tag{8}$$

be such that the sets

$$\left\{ \left(g_1^{(i)}, [\mathfrak{o}(g_2^{(i)})] \right) \right\}_{i=1}^{h}$$

partition $\mathcal{G}_n^a \times \mathcal{G}_m^a$. We will call this partition $\mathcal{P}_{n,m}$, where we suppress dependence on $\mathbf{g}$ and $\mathfrak{o}$ for ease of notation. We will define a Bayes optimal scheme Φ^* (independent of the choice of $\mathbf{g}$) piecewise on each element of this partition, and we will prove in Theorem 17 that Φ^* is level-k Bayes optimal for all $k \in [m-1]$:

$$\Phi^* \left(g_1^{(i)}, \mathfrak{o}(g_2^{(i)}), v^* \right)[1] \in \underset{u \in W}{\operatorname{argmax}} \ \mathbb{P}\left[\left(g_1^{(i)}, [\mathfrak{o}(g_2^{(i)})] \right)_{u=\mathfrak{o}(v^*)} \ \Big| \ \left(g_1^{(i)}, [\mathfrak{o}(g_2^{(i)})] \right) \right]$$

$$\Phi^* \left(g_1^{(i)}, \mathfrak{o}(g_2^{(i)}), v^* \right)[2] \in \underset{u \in W \setminus \{\Phi^*[1]\}}{\operatorname{argmax}} \ \mathbb{P}\left[\left(g_1^{(i)}, [\mathfrak{o}(g_2^{(i)})] \right)_{u=\mathfrak{o}(v^*)} \ \Big| \ \left(g_1^{(i)}, [\mathfrak{o}(g_2^{(i)})] \right) \right]$$

$$\vdots$$

$$\Phi^* \left(g_1^{(i)}, \mathfrak{o}(g_2^{(i)}), v^* \right)[m] \in \underset{u \in W \setminus \{\cup_{j \in [m-1]} \Phi^*[j]\}}{\operatorname{argmax}} \ \mathbb{P}\left[\left(g_1^{(i)}, [\mathfrak{o}(g_2^{(i)})] \right)_{u=\mathfrak{o}(v^*)} \ \Big| \ \left(g_1^{(i)}, [\mathfrak{o}(g_2^{(i)})] \right) \right], \tag{9}$$

with ties broken arbitrarily but deterministically. We refer the interested reader to Appendix B.1 for discussion of the case where ties are allowed in the ranking function. For each element

$$(g_1, g_2) \in \left(g_1^{(i)}, [\mathfrak{o}(g_2^{(i)})] \right) \setminus \left\{ \left(g_1^{(i)}, g_2^{(i)} \right) \right\},$$

choose the permutation σ such that $\mathfrak{o}(g_2) = \sigma(\mathfrak{o}(g_2^{(i)}))$, and define

$$\Phi^*(g_1^{(i)}, \mathfrak{o}(g_2), v^*) = \sigma(\Phi^*(g_1^{(i)}, \mathfrak{o}(g_2^{(i)}), v^*)).$$

Lastly, the following theorem shows that this scheme (uniquely defined up to tie-breaking) is indeed Bayes optimal. A proof can be found in Appendix A.1.

Theorem 17 *Let $\mathfrak{o} \in \mathfrak{D}_W$ be an obfuscating function, and let*

$$\mathbf{g} = \left\{ \left(g_1^{(i)}, g_2^{(i)} \right) \right\}_{i=1}^{h}$$

be such that the sets

$$\left\{ \left(g_1^{(i)}, [\mathfrak{o}(g_2^{(i)})] \right) \right\}_{i=1}^{h}$$

16

partition $\mathcal{G}_n^a \times \mathcal{G}_m^a$. Let $\Phi^* = \Phi_{\mathbf{g}}^*$ be as defined in Equation (9). Suppose that $(G_1, G_2) \sim F_{c,\theta}^{(n,m)} \in \mathcal{N}_{n,m}$ with $F_{c,\theta}^{(n,m)}$ supported on $\mathcal{G}_n^a \times \mathcal{G}_m^a$, and consider a vertex of interest $v^* \in C$. We have that $L_k(\Phi^*, v^*) = L_k^*(v^*)$ for all $k \in [m-1]$, partitions $\mathbf{g}$, and all obfuscating functions $\mathfrak{o}$.

Remark 18 The effect of symmetries on Theorem 23 is both subtle and cumbersome, as the specific tie-breaking procedures used in the analogue of Eq. (9) are of great import. To this end, consider $\mathbf{g}$ to be defined as above, and let $T \in \mathcal{T}_W$ be the ordering that specifies the (fixed but otherwise arbitrary) scheme by which elements within each $\mathcal{I}(v; \mathfrak{o}(g_2))$ are ordered. Informally, we will first rank the sets $\mathcal{I}(v; \mathfrak{o}(g_2))$ (rather than the individual vertices), and then use T to rank within and across each of the $\mathcal{I}(v; \mathfrak{o}(g_2))$. Full detail is provided below.

For each $w \in W$ and $v \in V(g_2)$, define

$$
(g_1, [\mathfrak{o}(g_2)])_{\mathcal{I}(w; \mathfrak{o}(g_2)) = \mathfrak{o}(v)} = \Big\{ (g_1, \tilde{g}_2) \text{ s.t. } \exists \text{ iso. } \sigma \text{ with } \mathfrak{o}(\tilde{g}_2) = \sigma(\mathfrak{o}(g_2)),
$$
$$
\text{and } \sigma(u) = \mathfrak{o}(v) \text{ for some } u \in \mathcal{I}(w; \mathfrak{o}(g_2)) \Big\}. \tag{10}
$$

As above, we will define the Bayes' optimal VN scheme on each element of the partition provided via $\mathbf{g}$. We first define a ranking Ψ of the sets

$$
\left\{ \left(g_1^{(i)}, [\mathfrak{o}(g_2^{(i)})] \right)_{\mathcal{I}(u; \mathfrak{o}(g_2^{(i)})) = \mathfrak{o}(v^*)} \right\}
$$

indexed by u, and then will use T to give the total ordering from Ψ. To wit, for each $i \in [k]$, iteratively define (where ties in the argmax are broken in an arbitrary but nonrandom manner)

$$
\Psi\left(g_1^{(i)}, \mathfrak{o}(g_2^{(i)}), v^* \right)[1] \in \underset{\mathcal{I}(u; \mathfrak{o}(g_2^{(i)})) \subset W}{\operatorname{argmax}} \mathbb{P}\left[\left(g_1^{(i)}, [\mathfrak{o}(g_2^{(i)})] \right)_{\mathcal{I}(u; \mathfrak{o}(g_2^{(i)})) = \mathfrak{o}(v^*)} \,\middle|\, \left(g_1^{(i)}, [\mathfrak{o}(g_2^{(i)})] \right) \right]
$$

$$
\Psi\left(g_1^{(i)}, \mathfrak{o}(g_2^{(i)}), v^* \right)[2] \in \underset{\mathcal{I}(u; \mathfrak{o}(g_2^{(i)})) \subset W \setminus \{\Psi[1]\}}{\operatorname{argmax}} \mathbb{P}\left[\left(g_1^{(i)}, [\mathfrak{o}(g_2^{(i)})] \right)_{\mathcal{I}(u; \mathfrak{o}(g_2^{(i)})) = \mathfrak{o}(v^*)} \,\middle|\, \left(g_1^{(i)}, [\mathfrak{o}(g_2^{(i)})] \right) \right]
$$

$$
\vdots
$$

$$
\Psi\left(g_1^{(i)}, \mathfrak{o}(g_2^{(i)}), v^* \right)[\ell] \in \underset{\mathcal{I}_{\mathfrak{o}(g_2)}(u) \subset W \setminus \{\cup_{j=1}^{\ell-1} \Psi[j]\}}{\operatorname{argmax}} \mathbb{P}\left[\left(g_1^{(i)}, [\mathfrak{o}(g_2^{(i)})] \right)_{\mathcal{I}(u; \mathfrak{o}(g_2^{(i)})) = \mathfrak{o}(v^*)} \,\middle|\, \left(g_1^{(i)}, [\mathfrak{o}(g_2^{(i)})] \right) \right]. \tag{11}
$$

For each element

$$
(g_1, g_2) \in (g_1^{(i)}, [\mathfrak{o}(g_2^{(i)})]) \setminus \{(g_1^{(i)}, g_2^{(i)})\},
$$

choose an isomorphism σ such that $\mathfrak{o}(g_2) = \sigma(\mathfrak{o}(g_2^{(i)}))$, and define

$$
\Psi(g_1, \mathfrak{o}(g_2), v^*) = \sigma(\Psi(g_1^{(i)}, \mathfrak{o}(g_2^{(i)}), v^*)).
$$

Note that the choice of isomorphism σ does not impact the definition of Ψ.

For each $(g_1, g_2) \in \mathcal{G}_n \times \mathcal{G}_m$, we define a VN scheme Φ_T^* from Ψ as follows:
1. Initialize $\Phi_T^*(g_1, \mathfrak{o}(g_2), v^*)$ as an empty list; initialize $j = 1$;
2. If $\Psi(g_1, \mathfrak{o}(g_2), v^*)[j]$ is nonempty, add the top ranked (according to T) element from $\Psi(g_1, \mathfrak{o}(g_2), v^*)[j]$ to the end of $\Phi_T^*(g_1, \mathfrak{o}(g_2), v^*)$, else do nothing; set $j = j + 1 \pmod{|\Psi(g_1, \mathfrak{o}(g_2), v^*)|}$
3. Repeat Step 2 until there are no more vertices to add to $\Phi_T^*(g_1, \mathfrak{o}(g_2), v^*)$.
If $T[1] = \mathfrak{o}(v^*)$, then $\Phi_T^*(g_1, \mathfrak{o}(g_2), v^*)$ is Bayes optimal (as in Theorem 17) in the sense of Definition 16. See Appendix A.1 for details.

Example 1, continued. Let $(G_1, G_2) \sim R$-$ER(P)$ for $P, R \in \mathbb{R}^{n \times n}$. Under mild model assumptions, we have that $\lim_{n \to \infty} L_k^*(v^*) = 0$ for any fixed k. This is due to the fact that the optimal graph matching of G_1 to $\mathfrak{o}(G_2)$ will almost surely recover the true vertex labels of $\mathfrak{o}(G_2)$ for n suitably large; i.e.,

$$\text{argmin}_{Q \in \Pi_n} \|AQ - QB\|_F = \{I_n\} \text{ with probability} \to 1,$$

where Π_n is the set of $n \times n$ permutation matrices, A is the adjacency matrix for G_1 and B the adjacency matrix for G_2. More concretely, we have the following theorem adapted to our present setting from Lyzinski and Sussman (2018). A proof sketch can be found in Appendix B.

Theorem 19 *Let $(A, B) \sim R$-ER(P), and for any fixed permutation matrix Q define the random variable $\delta(Q) := \|AQ - QB\|_F$. Define $0 < \epsilon := \min_{i,j; i \neq j} 2R_{i,j} P_{i,j}(1 - P_{i,j})$. There exist positive constants c_1, c_2 such that if $\epsilon^2 > c_1 \log(n)/n$, then for sufficiently large n,*

$$\mathbb{P}\left(\exists\, Q \in \Pi_n \setminus \{I_n\} \text{ s.t. } \delta(Q) \leq \delta(I_n)\right) \leq 2\exp\left\{-c_2 \epsilon^2 n\right\}.$$

Similarly to the Bayes optimal scheme, we define the graph matching VN scheme, denoted Φ_M, separately on each element of $\mathcal{P}_{n,n}$. For a given $(\tilde{g}_1, [\mathfrak{o}(\tilde{g}_2)]) \in \mathcal{P}_{n,n}$, let (g_1, g_2) be an fixed element in $(\tilde{g}_1, [\mathfrak{o}(\tilde{g}_2)])$. Define $\Phi_M(g_1, \mathfrak{o}(g_2), v^*)[1]$ to be a fixed but arbitrary element from

$$\left\{\mathfrak{r}^{-1}(v^*) \text{ s.t. } Q_{\mathfrak{r}} \in \text{argmin}_{Q \in \Pi_n} \|AQ - QB\|_F\right\}, \tag{12}$$

where each $\mathfrak{r} : W \to V_2$ appearing above is a bijection and $Q_{\mathfrak{r}}$ its associated permutation matrix (having identified both W and V_2 with the set $[n]$). Define

$$\mathfrak{R}_1 = \left\{\mathfrak{r} : W \to V_2 \text{ s.t. } \mathfrak{r} \text{ is a bijection with } \mathfrak{r}^{-1}(v^*) = \Phi_M(g_1, \mathfrak{o}(g_2), v^*)[1]\right\}.$$

If $j > 1$, define $\Phi_M(g_1, \mathfrak{o}(g_2), v^*)[j]$ to be any element of

$$\left\{\mathfrak{r}^{-1}(v^*) \text{ s.t. } Q_{\mathfrak{r}} \in \underset{Q \in \Pi(n) \setminus \{Q_{\mathfrak{r}} : \mathfrak{r} \in \bigcup_{\ell=1}^{j-1} \mathfrak{R}_\ell\}}{\text{argmin}} \|AQ - QB\|_F\right\},$$

where $\mathfrak{R}_j$ is defined analogously to $\mathfrak{R}_1$. For each element $(g_1', g_2') \in (\tilde{g}_1, [\mathfrak{o}(\tilde{g}_2)]) \setminus \{(g_1, g_2)\}$, choose a permutation σ such that $\mathfrak{o}(g_2') = \sigma(\mathfrak{o}(g_2))$, and we then define $\Phi_M(g_1, \mathfrak{o}(g_2'), v^*) = \sigma(\Phi_M(g_1, \mathfrak{o}(g_2), v^*))$. Theorem 19 states that under mild model assumptions, we have that $\Phi_M(G_1, \mathfrak{o}(G_2), v^*)[1] = \mathfrak{o}(v^*)$ asymptotically almost surely, and thus $\lim_{n \to \infty} L_1(\Phi_M, v^*) = 0$. Indeed, in this setting, for any fixed $k \geq 1$, $\lim_{n \to \infty} L_k(\Phi_M, v^*) = 0$. It is then immediate

that $\lim_{n\to\infty} L_k^*(v^*) = 0$ in this model for any fixed k.

The next two examples serve to illustrate how the level-k Bayes error behaves in the presence of stochastically indistinguishable vertices. In essence, we cannot hope to perform better than randomly ordering stochastically equivalent vertices.

Example 3, continued. Let G_1 and G_2 be independent $ER(n,p)$ graphs. Since the vertices are stochastically indistinguishable within each of the two graphs, no nomination scheme can do better than random chance in this model. Thus, with $c = n$, we have that $L_k^*(v^*) = (1 - k/n)(1 + o(1))$ for all $k \in [n-1]$ and all $v^* \in [n]$.

Example 4 Let $p_1, p_2, q \in [0,1]$ with $1 \ge p_1 > p_2 \ge 0$ and $q \ne p_1, p_2$. Define the matrix

$$B = \begin{pmatrix} p_1 & q \\ q & p_2 \end{pmatrix},$$

and let G_1, G_2 be independent $SBM(2, B, b_n)$ graphs where $b_n(i) = 1$ if $i \le n/2$ (n even) and $b_n(i) = 2$ if $i > n/2$. With $c = n$ and the correspondence equal to the identity function, let $(k_n)_{n=2}^\infty$ be a nondecreasing divergent sequence satisfying $k_n \le n/2$ for all $n > 1$, then $\lim_n L_{k_n}^*(v^*) = \lim_n [(1 - 2k_n/n) \vee 0]$ for all v^*. Indeed, $L_{k_n}^*$ is asymptotically equivalent to randomly ordering the $n/2$ vertices in G_2 that are stochastically equivalent to v^*.

4. VN Consistency

With the definition of Bayes optimality and the Bayes optimal scheme in hand, it is now possible to define a notion of consistent vertex nomination analogous to consistent classification in the pattern recognition literature. Before defining a consistent VN rule (i.e., a sequence of VN schemes), we must first define the notion of sequences of nominatable distributions with *nested cores*. Such sequences of distributions are necessary in order to speak sensibly of a sequence of vertex nomination problem instances.

Definition 20 *Let* $\mathbf{F} = \left(F_{c_n,\theta_n}^{(n,m_n)}\right)_{n=n_0}^\infty$ *be a sequence of distributions such that* $F_{c_n,\theta_n}^{(n,m_n)} \in \mathcal{N}_{n,m_n}$ *for all* n. *We say that* $\mathbf{F}$ *has* nested cores *if for all* $n_0 \le n_1 < n_2$, *if*

$$(G_1, G_2) \sim F_{n_1} = F_{c_{n_1},\theta_{n_1}}^{(n_1,m_{n_1})} \text{ and } (\widetilde{G}_1, \widetilde{G}_2) \sim F_{n_2} = F_{c_{n_2},\theta_{n_2}}^{(n_2,m_{n_2})},$$

we have, letting C_1 *and* C_2 *be the core vertices associated with* F_{n_1} *and* F_{n_2} *respectively, and denoting the junk vertices* $J_{1,1}, J_{2,1}, J_{1,2}, \widetilde{J}_{2,2}$ *analogously,*
[i.] $V(G_1) = C_1 \cup J_{1,1} \subset V(\widetilde{G}_1) = C_2 \cup J_{2,1}$;
[ii.] $V(G_2) = C_2 \cup J_{1,2} \subset V(\widetilde{G}_2) = C_2 \cup J_{2,2}$;
[iii.] $C_1 \subset C_2$.

We are now ready to define a consistent VN rule.

Definition 21 *Let*

$$\mathbf{F} = \left(F_n = F_{c_n,\theta_n}^{(n,m_n)}\right)_{n=n_0}^\infty$$

be a sequence of distributions such that $F_n \in \mathcal{N}_{n,m(n)}$ with nested cores satisfying $\lim_{n \to \infty} m_n = \infty$. For a given non-decreasing sequence (k_n), we say that a VN rule $\mathbf{\Phi} = (\Phi_{n,m_n})_{n=n_0}^{\infty}$ is level-(k_n) consistent at v^* with respect to $\mathbf{F}$ if

$$\lim_{n \to \infty} L_{k_n}(\Phi_{n,m_n}, v^*) - L_{k_n}^*(v^*) = 0,$$

for any sequence of obfuscating functions of V_2 with $|V_2| = m_n$. If a rule $\mathbf{\Phi}$ is level-(k_n) consistent at v^* for a constant sequence $k_n = k$, $n = 1, 2, \ldots$, then we say simply that $\mathbf{\Phi}$ is level-k consistent.

Remark 22 Equation (5) has an interesting implication for VN consistency in the setting where $L_{k_n}^*(v^*) \to 0$. In this case, level-(k_n) consistency of a VN rule $\mathbf{\Phi}$ implies that $\mathbf{\Phi}$ is (k_n')-consistent for all (k_n') such that $\liminf \frac{k_n'}{k_n} \geq 1$. We conjecture that this implication holds true for the case where $L_{k_n}^*(v^*) \to c > 0$, but this problem remains open at present.

Example 1, continued. Let $\mathbf{F} = (F_{n,\theta_n=(P_n,R_n)}^{(n,n)})$ be a sequence of R_n-ER(P_n) random graph models in $\mathcal{N}_{n,n}$ for some sequence of probability matrices $(P_n)_{n=n_0}^{\infty}$ and correlation matrices $(R_n)_{n=n_0}^{\infty}$. Under mild model assumptions (see Theorem 19), the graph matching vertex nomination rule $\mathbf{\Phi}_M$ defined in Equation (12) above is level-1 consistent, and hence level-(k_n) consistent for all (k_n) sequences.

Example 3, continued. Let $\mathbf{F} = (F_{n,\theta_n=p}^{(n,n)})$ be a sequence of independent ER(n, p) random graph models in $\mathcal{N}$. All vertex nomination rules are level-(k_n) consistent for all (k_n) sequences. This holds for all possible values of $c \in [n]$ in the nested sequence of ER(n, p) distributions, as all VN rules have effectively chance performance, regardless of core size under this model.

We define the consistency of a VN rule with respect to a broad class of graph sequences, and it is perhaps no surprise that there cannot be any level-(k_n) universally consistent VN rules. Indeed, even for constant sequences $k_n := k$ (i.e., those that are level-(k_n) consistent for all sequences of nominatable distributions $\mathbf{F}$ with nested cores) there cannot be any level-k universally consistent VN rules. To prove this result, we will first establish an analogue to the "arbitrary poor performance" theorems for classifiers (see, e.g., Devroye et al., 1997, Theorem 7.1), which state that for a fixed n and m, any VN scheme can be shown to have arbitrarily poor performance with respect to a well-chosen adversarial distribution $F_{c,\theta}^{(n,m)}$. Our theorem mirrors the classical classification literature, as for a given classification rule, there exists "a sufficiently complex distribution for which the sample size n is hopelessly small" (Devroye et al., 1997, pg. 111), so that a classification rule can be made to perform arbitrarily poorly by selecting a suitably complex data distribution. Nonetheless, in the case of classification, this model complexity and the implicit dependence on n can be overcome asymptotically by a classification rule. That is, universally consistent classifiers exist (see, for example, Stone, 1977; Steinwart, 2002; Tang et al., 2013). In contrast, in the VN problem, the complexity of the model generating the data can also grow in n, which effectively thwarts the ability of a VN rule to asymptotically overcome a sequence of adversarial graph models. Formalizing the above, we arrive at the following theorem, a proof of which can be found in Appendix A.2.

Theorem 23 *Let n and m be large enough to guarantee the existence of asymmetric graphs $g_1 \in \mathcal{G}_n$ and $g_2 \in \mathcal{G}_m$. Consider a VN scheme $\Phi \in \mathcal{V}_{n,m}$, obfuscating function $\mathfrak{o}$, and strictly increasing sequence $(\epsilon_i)_{i=1}^m$ satisfying $\epsilon_i \in (0, \frac{i}{m})$. Then there exists a distribution $F_{c,\theta}^{(n,m)} \in \mathcal{N}_{n,m}$ over $\mathcal{G}_n \times \mathcal{G}_m$ and $v^* \in C$ such that for each $k \in [c-1]$,*

$$L_k^*(v^*) \leq \epsilon_{m-k} < 1 - \frac{k}{m} < 1 - \epsilon_k < L_k(\Phi, v^*),$$

where $1 - \frac{k}{m}$ represents the error probability of chance performance, i.e., the error probability of a VN scheme in the independent Erdős-Rényi setting.

In the remainder of the section, we will suppress the dependence of $m = m_n$ on n. If we consider sequences $(\epsilon_{m,i})_{i=1}^m$ satisfying the assumptions of Theorem 23 and $\lim_n \epsilon_{m,m-k_n} = \epsilon \in (0,1)$ for a given (k_n) satisfying $k_n = o(m)$, we arrive at the following Corollary, namely that level-(k_n) universally consistent VN schemes do not exist for any sequence $(k_n)_{n=n_0}^\infty$ that does not grow as fast as $m = |V(G_2)|$.

Corollary 24 *Let $\epsilon \in (0,1)$ be arbitrary, and consider a VN rule $\mathbf{\Phi} = (\Phi_{n,m})$. For any nondecreasing sequence $(k_n)_{n=n_0}^\infty$ satisfying $k_n = o(m)$, there exists a sequence of distributions $(F_{c,\theta}^{(n,m)} \in \mathcal{N}_{n,m})_{n=n_0}^\infty$ with nested cores such that*

$$\limsup_{n \to \infty} L_{k_n}^*(v^*) = \epsilon < 1 = \lim_{n \to \infty} L_{k_n}(\Phi_{n,m}, v^*).$$

Corollary 24 has a number of practical implications. Below, we will briefly outline two such implications. Unlike in the classification setting, where universally consistent rules (e.g., k-nearest neighbors) are theoretically guaranteed to perform well in big-data settings, the VN practitioner enjoys no such certainty. Indeed, in VN, the practitioner first needs to identify the consistency class of a VN rule (i.e., the set of models for which the VN rule is consistent) before applying it in real settings. Unfortunately, identifying and enumerating these consistency classes is theoretically and practically nontrivial, and we are investigating theory and heuristics for this at present. In a streaming data setting, the performance of a universally consistent classifier will approach Bayes optimality for the distribution governing the data, and the classifier will be guaranteed to successfully adapt itself to any changes in the underlying data distribution. The lack of universal consistency in the VN setting implies that his is not the case for graph data, as the performance, as the performance of a consistent VN scheme in the streaming setting could precipitously decline in the presence of distributional shifts in the data. Recognizing these shifts and their potential impact on VN performance is paramount and is the subject of current research.

4.1. Global Consistency

We have just seen that no universally consistent VN schemes exist. This is a consequence of the complexity of the models available when choosing a sequence of nominatable distributions. Indeed, nested-core nominatable sequences $\mathbf{F}$ allow for (nearly) arbitrary dependence structure and model complexity as n increases: corresponding vertex behavior may be correlated (see Example 1), independent (see Example 4), or negatively correlated (see Example 5) across networks. This model flexibility is in service of modeling the complexity of real

world networks, but, as we will demonstrate below, restricting our model class to simpler dependency structures still does not necessarily guarantee the existence of universally consistent schemes.

It is thus natural to explore a weaker notion of consistency, namely consistency for a sufficiently large family of nominatable sequences rather than for all nominatable sequences.

Definition 25 *Let*
$$\mathfrak{F} = \left\{ \mathbf{F}_\alpha = \left(F_{c_n,\theta_n}^{(n,m_n),\alpha} \right)_{n=n_0}^\infty : \alpha \in \mathcal{A} \right\}$$

be a family of nominatable sequences, indexed by some set $\mathcal{A}$. We say that VN scheme Φ is level-(k_n) $\mathfrak{F}$-globally consistent if Φ is level-(k_n) consistent for every $\mathbf{F} \in \mathfrak{F}$. We call such a family level-(k_n) globally consistent.

The question of the maximal family $\mathfrak{F}$ for which a level-(k_n) $\mathfrak{F}$-globally consistent rule exists is of prime interest. While we cannot offer a satisfactory complete answer to that question in the present work, we do offer some examples of globally consistent families.

Example 1 continued: In settings where corresponding vertices have correlated neighborhood structures across networks, there is hope for finding globally consistent rules. In the ongoing Example 1, we have seen a simple example of this in the R-ER(P) model, in which the matrix of correlations R encodes a correspondence across the two graphs. As mentioned previously, Theorem 19 asserts that under some mild model assumptions on R and P in the R-ER(P) model, level-1 globally consistent VN rules exist (namely the graph matching VN scheme). If $\mathfrak{F}$ denotes the set of distributions obeying these model assumptions, then we have that level-(k_n) $\mathfrak{F}$-globally consistent rules exist for all sequences (k_n). While we do not expect the conditions of Theorem 19 to produce a maximal level-(k_n) globally consistent family for any given (k_n), this example nonetheless provides an important intuition for the properties such maximal families might possess.

Example 4 continued: The SBM provides a prime example of global consistency. Working in the one-graph framework of Remark 13, under appropriate growth conditions on the parameters of every sequence in family $\mathfrak{F}$, Theorem 6 in Lyzinski et al. (2016) implies the existence of a likelihood-based nomination scheme that is level-$(|U_1|)$ globally consistent for this family of models. Under similar growth conditions, Theorem 6 in Lyzinski et al. (2014) implies the existence of a level-$(|U_1|)$ globally consistent scheme based on spectral clustering, in which vertices are nominated based on their proximity to the vertex or vertices of interest.

Remark 26 An attempt at systematically constructing a maximal globally consistent family might begin by putting model restrictions onto elements of $\mathcal{N}_{n,m}$. A natural restriction to consider would be to demand that the models in $\mathfrak{F}$ be nested in the following sense: For $\mathbf{F} \in \mathfrak{F}$, if $(G_1, G_2) \sim F_{c_{n_2},\theta_{n_2}}^{(n_2,m_{n_2})}$ with $n_1 < n_2$, then $(G_1[[n_1]], G_2[[m_{n_1}]]) \stackrel{\mathcal{L}}{=} (G_1', G_2')$ where $(G_1', G_2') \sim F_{c_{n_1},\theta_{n_1}}^{(n_1,m_{n_1})}$. This property would allow us to consider "streaming" network models $\mathbf{F}$, where for $n_1 < n_2$, if (g_1, g_2) is realized from $(G_1, G_2) \sim F_{c_{n_2},\theta_{n_2}}^{(n_2,m_{n_2})}$, and (g_1', g_2') is realized from $(G_1', G_2') \sim F_{c_{n_1},\theta_{n_1}}^{(n_1,m_{n_1})}$ then (g_1, g_2) can be constructed by appropriately adding

$n_2 - n_1$ vertices to (g'_1, g'_2). Additionally, this would serve to mimic the nested nature of the data in the classification consistency literature. However, as we will see in Example 5, global consistency depends both on the dependency structure within each graph (as seen in Theorem 23) *and* the vertex correspondence (i.e., the potential dependency structure across graphs) encoded in the model.

4.2. Behavioral (In)consistency and Global (In)consistency

We suspect that if the vertices of interest have a common distinguishing probabilistic and/or topological characteristic (e.g., correlated neighborhoods, common SBM block structure, high network centrality, etc.) then a globally consistent rule may exist. Indeed, under mild model assumptions, this is the case in the R-ER(P) of Example 1; in the i.i.d. SBM of Example 4 where the correspondence is the identity function (Lyzinski et al., 2014); and in the i.i.d. ER of Example 3, to name a few. In each of these examples, there is a stochastic/topological similarity (or in the ER case, uniformity) between corresponding vertices across networks. In each, corresponding vertices behave similarly across networks. While we suspect that this behavioral similarity is not sufficient for global consistency, Example 5 demonstrates that behavioral inconsistencies within a family of nominatable distributions can preclude the existence of globally consistent nomination rules.

Example 5 For each n, consider n-vertex random graphs $G_1 \sim$ asym-SBM$(2, B_1, b_n^{(1)})$ independent of $G_2 \sim$ asym-SBM$(2, B_2, b_n^{(2)})$, where asym-SBM denotes the stochastic blockmodel distribution restricted to have support on asymmetric graphs. This restriction is made to avoid the unpleasantries of symmetries, and is asymptotically negligible as the SBMs considered in this example are asymptotically almost surely asymmetric.
Case 1. In this case, corresponding vertices behave similarly across networks. To wit, let $\mathbf{F} = (F_n)_{n=n_0}^{\infty}$ be the sequence of models where

$$B_1 = B_2 = \begin{pmatrix} p_1 & q \\ q & p_2 \end{pmatrix}, \ b_n^{(1)}(i) = b_n^{(2)}(i) = \begin{cases} 1 & \text{if } i \leq n/2; \\ 2 & \text{if } i > n/2, \end{cases}$$

$p_1 \neq p_2$, $c = n$, and the correspondence is the identity function. As stated before, in this model $L_{n/2}^*(v^*) \to 0$ for all $v^* \in C$. Without loss of generality, consider $v^* = v_1 = u_1 = 1$. If $\mathbf{\Phi}$ is consistent with respect to $\mathbf{F}$ then

$$\mathbb{P}_{F_n}(\text{rank}_{\Phi(G_1, \mathfrak{o}(G_2), v_1)}(\mathfrak{o}(u_1)) \geq n/2 + 1) \to 0.$$

By the distributional equivalence of vertices within the same block, and the consistency property in the definition of a VN scheme, for any $u, v \in b_n^{-1}(1)$, $k \in [n]$ we have that

$$\mathbb{P}_{F_n}(\text{rank}_{\Phi(G_1, \mathfrak{o}(G_2), v_1)}(\mathfrak{o}(u)) = k) = \mathbb{P}_{F_n}(\text{rank}_{\Phi(G_1, \mathfrak{o}(G_2), v_1)}(\mathfrak{o}(v)) = k).$$

Letting this common value be set to $\alpha_{k,n}$ (with $\beta_{k,n}$ defined similarly as the common value of $\mathbb{P}_{F_n}(\text{rank}_{\Phi(G_1, \mathfrak{o}(G_2), v_1)}(\mathfrak{o}(u)) = k)$ for u in block 2), we have that

$$\sum_{i=1}^{n} \mathbb{P}_{F_n}(\text{rank}_{\Phi(G_1, \mathfrak{o}(G_2), v_1)}(\mathfrak{o}(u_i)) = k) = 1 = \frac{n(\alpha_{k,n} + \beta_{k,n})}{2}$$

giving us that $\alpha_{k,n} = 2/n - \beta_{k,n}$. Consistency implies that

$$\sum_{k=1}^{n/2} \alpha_{k,n} \to 1,$$

which implies that

$$\sum_{k=1}^{n/2} \alpha_{k,n} = \sum_{k=1}^{n/2} (2/n - \beta_{k,n}) = 1 - \sum_{k=1}^{n/2} \beta_{k,n} \to 1,$$

implying $\sum_{k=1}^{n/2} \beta_{k,n} \to 0$. Therefore, for any u in block 2,

$$\mathbb{P}_{F_n}(\text{rank}_{\Phi(G_1, \mathfrak{o}(G_2), v_1)}(\mathfrak{o}(u)) \geq n/2 + 1) \to 1.$$

Case 2. In this case, corresponding vertices behave differently across networks. To wit, let $\tilde{\mathbf{F}} = (\tilde{F}_n)_{n=n_0}^{\infty}$ be the sequence of models where

$$B_1 = \begin{pmatrix} p_1 & q \\ q & p_2 \end{pmatrix}, \quad B_2 = \begin{pmatrix} p_2 & q \\ q & p_1 \end{pmatrix}, \quad b_n^{(1)}(i) = b_n^{(2)}(i) = \begin{cases} 1 \text{ if } i \leq n/2; \\ 2 \text{ if } i > n/2, \end{cases}$$

$c = n$, and the correspondence is the identity function. As in Case 1 considered above, in this model $L_{n/2}^*(v^*) \to 0$ for all $v^* \in C$, and, as above, consider $v^* = v_1 = u_1$. If $\boldsymbol{\Phi}$ is consistent with respect to $\tilde{\mathbf{F}}$ then

$$\mathbb{P}_{\tilde{F}_n}(\text{rank}_{\Phi(G_1, \mathfrak{o}(G_2), v_1)}(\mathfrak{o}(u_1)) \geq n/2 + 1) \to 0.$$

Note that if σ is the permutation such that

$$\sigma(i) = \begin{cases} i + n/2 & \text{if } i \leq n/2; \\ i - n/2 & \text{if } i > n/2, \end{cases}$$

then $\mathbb{P}_{F_n}(g_1, g_2) = \mathbb{P}_{\tilde{F}_n}(g_1, \sigma(g_2))$. Define

$$E_n = \{(g_1, g_2) \text{ s.t. } \text{rank}_{\Phi(g_1, \mathfrak{o}(g_2), v_1)}(\mathfrak{o}(u_1)) \geq n/2 + 1\}\}$$

i.e., and $\tilde{E}_n = \{(g_1, g_2) \text{ s.t. } (g_1, \sigma(g_2)) \in E_n\}$, i.e.,

$$\tilde{E}_n = \{(g_1, g_2) \text{ s.t. } \text{rank}_{\Phi(g_1, \mathfrak{o}(g_2), v_1)}(\mathfrak{o}(u_{n/2+1})) \geq n/2 + 1\}.$$

If $\boldsymbol{\Phi}$ is consistent with respect to $\mathbf{F}$ we have that $\mathbb{P}_{F_n}(E_n) \to 0$ which implies (as $(g_1, g_2) \in E_n \Leftrightarrow (g_1, \sigma(g_2)) \in \tilde{E}_n$) $\mathbb{P}_{\tilde{F}_n}(\tilde{E}_n) \to 0$. If $\boldsymbol{\Phi}$ is consistent with respect to $\tilde{\mathbf{F}}$ then $\mathbb{P}_{\tilde{F}_n}(E_n) \to 0$ and $\mathbb{P}_{\tilde{F}_n}(\tilde{E}_n) \to 1$. We arrive at a contradiction, and $\boldsymbol{\Phi}$ cannot be $(n/2)$-consistent with respect to both $\mathbf{F}$ and $\tilde{\mathbf{F}}$.

Although Example 5 may seem artificial, it is a simple representation of a common phenomenon observed in network data. Often the same entity can behave quite differently across networks; see Patsolic et al. (2017) for an example of this in social networks and Chen et al. (2016b) for an example of this in connectomics. In such a setting, intuition says that a universal scheme that works in both behavioral settings should not exist. Indeed, at least in the simple block model setting considered above, we see that no such scheme exists. Example 5 also highlights an important difference between the VN setting and the more standard classification framework. We already noted that classification's universal consistency relies on the distribution not changing in n, whereas in VN the distributions must vary with n (indeed, the graph sizes grow in n). Further, Example 5 shows that the nonexistence of a universally consistent scheme is not simply a consequence of changing the underlying distributional parameters with n, as these two SBM distributions are (essentially) fixed, in that the matrix B does not change with n. In this example the "training data" provided by G_1 cannot be uniformly beneficial for a single VN scheme across the two differing model settings we consider. Contrast this with the classification setting of Devroye et al. (1997), where the training data uniformly provides progressively better estimates of the class-conditional distributions. Indeed, the training data helps delineate potentially interesting vertices from non-interesting ones in G_2 in Case (1) for one VN scheme, and in Case (2) for another VN scheme, but there does not exist a VN scheme that achieves this desired class separation across both cases.

Remark 27 In the cases considered in Example 5, if we introduce positive edge-wise correlation of

$$\rho \leq \sqrt{\min\left(\frac{p_1(1-p_2)}{p_2(1-p_1)}, \frac{p_2(1-p_1)}{p_1(1-p_2)}\right)}$$

into both Case 1 and Case 2, then under mild assumptions on the growth of p_1 and p_2, joint consistency can be recovered via a USVT centered graph matching nomination scheme (see Lyzinski and Sussman, 2018, for details). This example demonstrates that it is sometimes possible to toggle a family of models to allow for global consistency. A note of caution is needed, however, as in this particular example the correlation ρ is introducing a behavioral consistency across networks that addresses the precise issue brought forth by the behavioral inconsistency in Example 5. In other, more nuanced model families, we do not expect the global-consistency modification (if it indeed exists) to be as straightforward as adding additional edge-correlation into the model.

4.3. Vertex Nomination on Networks with Node Covariates

It is natural to ask if incorporating vertex features into the VN framework can resolve the lack of universally consistent VN schemes. While straightforward to implement, the ameliorating effect of features is significantly more nuanced. Before defining the VN scheme with features, we need the following extension of $\mathcal{I}(v; g)$ for $g \in \mathcal{G}_n$ and $v \in V(g)$. Letting $\mathcal{X}$ be the space of vertex features for graphs in $\mathcal{G}_n$, for $g \in \mathcal{G}_n$, $v \in V(g)$, and $X \in \mathcal{X}^n$ we define

$$\widetilde{\mathcal{I}}(v; g, X) = \{u \in V(g) : \exists \text{ automorphism } \tau \text{ of } g \text{ s.t. } \tau(v) = u \text{ and } X_u = X_v\},$$

where X_v is the feature associated to v via X.

Definition 28 (Vertex Nomination (VN) Scheme with features) *Let $\mathcal{X}$ (resp., $\mathcal{Y}$) be the space of vertex features of graphs in $\mathcal{G}_n$ (resp., $\mathcal{G}_m$). For $n, m > 0$ fixed, obfuscating set W and obfuscating function $\mathfrak{o} \in \mathfrak{O}_W$,* a vertex nomination scheme with features *is a function*

$$\Phi : \mathcal{G}_n \times \mathcal{X}^n \times \mathfrak{o}(\mathcal{G}_m) \times \mathcal{Y}^m \times V_1 \to \mathcal{T}_W$$

satisfying the following consistency property: If for each $u \in V_2$, we define

$$rank_{\Phi(g_1, X, \mathfrak{o}(g_2), \mathfrak{o}(Y), v^*)}\big(\mathfrak{o}(u)\big)$$

to be the position of $\mathfrak{o}(u)$ in the total ordering provided by $\Phi(g_1, X, \mathfrak{o}(g_2), \mathfrak{o}(Y), v^*)$, and we define $\mathfrak{r}_\Phi : \mathcal{G}_n \times \mathcal{X}^n \times \mathcal{G}_m \times \mathcal{Y}^m \times \mathfrak{O}_W \times V_1 \times 2^{V_2} \mapsto 2^{[m]}$ via

$$\mathfrak{r}_\Phi(g_1, X, \mathfrak{o}(g_2), \mathfrak{o}(Y), v^*, S) = \{rank_{\Phi(g_1, X, \mathfrak{o}(g_2), \mathfrak{o}(Y), v^*)}\big(\mathfrak{o}(u)\big) \ s.t. \ u \in S\},$$

then we require that for any $g_1 \in \mathcal{G}_n$, $g_2 \in \mathcal{G}_m$, $v^* \in V_1$, $X \in \mathcal{X}^n$, $Y \in \mathcal{Y}^m$, obfuscating functions $\mathfrak{o}_1, \mathfrak{o}_2 \in \mathfrak{O}_W$ and any $u \in V(g_2)$,

$$\mathfrak{r}_\Phi\big(g_1, X, \mathfrak{o}_1(g_2), \mathfrak{o}_1(Y), v^*, \widetilde{\mathcal{I}}(u; g_2, Y)\big) = \mathfrak{r}_\Phi\big(g_1, X, \mathfrak{o}_2(g_2), \mathfrak{o}_2(Y), v^*, \widetilde{\mathcal{I}}(u; g_2, Y)\big). \quad (13)$$

We let $\widetilde{\mathcal{V}}_{n,m}$ denote the set of all such VN schemes.

It is immediate that if the features are sufficiently informative, consistency can be established with features where it could not be without. Indeed, consider in Example 5 features that encode the community memberships of a few vertices (e.g., a few vertices whose correspondences across the two graphs are known a priori). Combined with spectral methods, these would be sufficient for consistent VN under either behavior regime. It is also immediate that the fundamental idea presented in Example 5 has an analogue when vertex features are available, as illustrated by the following example.

Example 6 For each n, consider n-vertex random graphs $G_1 \sim$ asym-SBM$(3, B_1, b_n^{(1)})$ independent of $G_2 \sim$ asym-SBM$(3, B_2, b_n^{(2)})$, where asym-SBM again indicates the stochastic block model with support restricted to the asymmetric graphs.
Case 1. In this case, corresponding vertices behave similarly across networks. To wit, let $\mathbf{F} = (F_n)_{n=n_0}^\infty$ be the sequence of models where $3|n$ and

$$B_1 = B_2 = \begin{pmatrix} p_1 & q & q \\ q & p_2 & q \\ q & q & p_2 \end{pmatrix}, \ b_n^{(1)}(i) = b_n^{(2)}(i) = \begin{cases} 1 & \text{if } i \leq n/3; \\ 2 & \text{if } i \in (n/3, 2n/3] \\ 3 & \text{if } i > n/3, \end{cases}$$

$p_1 \neq p_2$, $c = n$, the correspondence is the identity function, and $v^* = 1$ is in block 1.
Case 2. In this case, corresponding vertices behave differently across networks. To wit, let $\mathbf{F} = (F_n)_{n=n_0}^\infty$ be the sequence of models where $3|n$ and

$$B_1 = \begin{pmatrix} p_1 & q & q \\ q & p_2 & q \\ q & q & p_2 \end{pmatrix}, \ B_2 = \begin{pmatrix} p_2 & q & q \\ q & p_1 & q \\ q & q & p_2 \end{pmatrix}, \ b_n^{(1)}(i) = b_n^{(2)}(i) = \begin{cases} 1 & \text{if } i \leq n/3; \\ 2 & \text{if } i \in (n/3, 2n/3] \\ 3 & \text{if } i > n/3, \end{cases}$$

$p_1 \neq p_2$, $c = n$, the correspondence is the identity function, and $v^* = 1$ is in block 1. Similar to Example 5, without features no VN scheme can be $(n/3)$-consistent for both Cases 1 and 2 simultaneously. In a similar fashion, if we consider features X and Y defined via

$$X_v = Y_v = \begin{cases} 1 & \text{if } b^{(1)}(v) = 1 \\ -1 & \text{if } b^{(1)}(v) = 2, 3, \end{cases}$$

then joint consistency is achievable for both Cases 1 and 2, for example by relying on features and ignoring graph structure. However, if we consider features X and Y defined via

$$X_v = Y_v = \begin{cases} 1 & \text{if } b^{(1)}(v) = 1, 2 \\ -1 & \text{if } b^{(1)}(v) = 3, \end{cases}$$

then joint consistency is again not achievable for both Cases 1 and 2 simultaneously.

This example demonstrates that features, in general, are not enough to ensure universal consistency. Nevertheless, insofar as features supply additional information, they have the potential to improve VN performance. A more thorough examination of the effect and effectiveness of vertex features in VN is beyond the scope of this work, and is the subject of current research.

5. Discussion

In this work, we have introduced a notion of consistency for the vertex nomination task that better reflects the broad range of models under which VN may be deployed. Rather than being restricted to the stochastic block model structure required in previous formulations of the problem, our framework allows for arbitrary dependence structure both within and between graph pairs, while encompassing the original SBM formulation of the problem. Additionally, we have demonstrated how this framework relates to the well-studied notion of Bayes optimal classifiers in the pattern recognition literature. Unlike in the classification setting, we have seen that while Bayes optimal VN schemes always exist, no universally consistent scheme exists. This fact is due essentially to the additional leeway provided by the graph model, in which observing more vertices does not necessarily correspond to receiving more information about the underlying distribution. This is in contrast to the classification setting studied in Stone (1977) and others (Devroye et al., 1997), in which observing more samples allows more accurate estimation of the underlying distribution and class boundary. For this reason, one especially interesting line of investigation concerns the nominatable distributions for which larger n does indeed correspond to more information about the underlying graph distribution. A simple example of this is the initial formulation of the vertex nomination problem, in which observing more vertices allows one to better estimate the model parameters, including the block memberships, and thus more accurately identify the vertices from the interesting block. We suspect that the essential property at play here is that under models of this sort, each vertex is analogous to a sample from a single distribution, though this may not be in and of itself a sufficient condition for consistency. For example, in the case of (G_1, G_2) being i.i.d. or ρ-correlated marginally identical draws from a random dot product graph model with the identity correspondence, each vertex (along with its incident edges) is, in a sense, a noisy sample from the underlying latent

27

position distribution. Hence, for large n, one can estimate the latent positions or their distribution to arbitrary accuracy, and provided that the latent positions of the interesting vertices are suitably separated from those of the rest of the graph, one should have VN consistency for the collection of these latent position models.

More broadly, it would be good to better understand whether there exist families of nominatable distributions $\mathfrak{F}$ for which certain VN schemes are consistent, and precisely how large these families can be made to be. In a similar vein, it would be of interest to explore how the dependence structure allowed both within and between graphs influences vertex nomination. In particular, if one rules out certain pathologically hard dependence structures as considered in Example 5, can one obtain global consistency with respect to this restricted set of distributions? We hope to explore these questions in future work.

We are also exploring alternative formulations of the VN problem and alternate formulations of the VN loss function. While the extension to multiple vertices of interest in each network G_1 and G_2 is straightforward, we are considering several generalizations of the VN problem considered here. One formulation of prime interest in applications (especially in connectomics and social networks) is as follows: given a collection of vertices of interest in one graph, find those that play a similar structural (based on the topology of the underlying network) or functional (based on vertex or edge covariates) role in the other graph. In addition, as seen in Section 4.3 the impact on VN consistency (and the potential existence of universally consistent schemes) when incorporating edge and vertex covariates into the VN framework is of prime interest, and a deeper analysis of the VN inference task in this framework is the subject of our current and future work.

The loss function considered in the present work is an analogue of the 0/1 recall-at-k loss function in the information retrieval literature. Under this loss function, we have shown that no universally consistent VN rule exists. It is natural to ask whether alternative loss functions can be considered under which universal consistency is achievable. While we conjecture that Example 5 will nearly always provide a counterexample to universal consistency, this question remains open and is the subject of current research.

Acknowledgments

Appendix A. Proofs of Main Results

Here we collect the proofs of the two main theorems in this work, Theorems 17 and 23.

A.1. Theorem 17 and Remark 18

In this section, we present proofs of Theorem 17 and the claim in Remark 18.

Proof [Proof of Theorem 17]

Let Φ be an arbitrary VN scheme and Φ^* the scheme defined at Eq. (9). Recall the definition

$$(g_1, [\mathfrak{o}(g_2)])_{w=\mathfrak{o}(v)} = \Big\{ (g_1, \tilde{g}_2) \text{ s.t. } \exists \text{ iso. } \sigma \text{ with } \mathfrak{o}(\tilde{g}_2) = \sigma(\mathfrak{o}(g_2)), \text{ and } \sigma(w) = \mathfrak{o}(v^*) \Big\}.$$

With $\mathbf{g}$ defined as in the theorem, note that for each $i \in [h]$,

$$
\begin{aligned}
U_{i,\mathbf{g}}^j :&= \Big\{ (g_1, g_2) \in \big(g_1^{(i)}, [\mathfrak{o}(g_2^{(i)})] \big) \text{ s.t. } \mathrm{rank}_{\Phi(g_1, \mathfrak{o}(g_2), v^*)}(\mathfrak{o}(v^*)) = j \Big\} \\
&= \Big\{ (g_1, g_2) \in \big(g_1^{(i)}, [\mathfrak{o}(g_2^{(i)})] \big) \text{ s.t. } \Phi(g_1, \mathfrak{o}(g_2), v^*)[j] = \mathfrak{o}(v^*) \Big\} \\
&= \Big\{ (g_1, g_2) \in \big(g_1^{(i)}, [\mathfrak{o}(g_2^{(i)})] \big) \text{ s.t. } \exists \text{ iso. } \sigma \text{ s.t. } \sigma(\mathfrak{o}(g_2^{(i)})) = \mathfrak{o}(g_2) \text{ and} \\
&\qquad\qquad \sigma(\Phi(g_1^{(i)}, \mathfrak{o}(g_2^{(i)}), v^*)[j]) = \mathfrak{o}(v^*) \Big\} \\
&= \big(g_1^{(i)}, [\mathfrak{o}(g_2^{(i)})] \big)_{\Phi(g_1^{(i)}, \mathfrak{o}(g_2^{(i)}), v^*)[j] = \mathfrak{o}(v^*)}.
\end{aligned}
$$

For each $i \in [h]$ define $p_{i,\Phi} \in [0,1]^m$ via

$$p_{i,\Phi}[j] := \mathbb{P}_{F_{c,\theta}^{(n,m)}} \Big(U_{i,\mathbf{g}}^j \,\big|\, (g_1^{(i)}, [\mathfrak{o}(g_2^{(i)})]) \Big).$$

Observe that for each $i \in [h]$, it is immediate that p_{i,Φ^*} majorizes $p_{i,\Phi}$. To see this, note that for any fixed $\xi \in [m]$, letting $q_{i,\Phi}^\xi$ be $(p_{i,\Phi}[j])_{j=1}^\xi$ with entries sorted in descending order, we have $p_{i,\Phi^*}[j] \geq q_{i,\Phi}^\xi[j]$ for all $j \in [\xi]$, and majorization follows immediately. With $\mathcal{P}_{\mathbf{g}}$ denoting the partition induced by $\mathbf{g}$, this majorization property implies

$$
\begin{aligned}
L_\xi(\Phi, v^*) &= 1 - \sum_{j=1}^\xi \mathbb{P}\big(\mathrm{rank}_{\Phi(G_1, \mathfrak{o}(G_2), v^*)}(\mathfrak{o}(v^*)) = j \big) \\
&= 1 - \sum_{\mathcal{P}_{\mathbf{g}}} \sum_{j=1}^\xi \mathbb{P}\big[U_{i,\mathbf{g}}^j \,\big|\, (g_1^{(i)}, [\mathfrak{o}(g_2^{(i)})]) \big] \, \mathbb{P}\big[(g_1^{(i)}, [\mathfrak{o}(g_2^{(i)})]) \big] \\
&= \sum_{\mathcal{P}_{\mathbf{g}}} \left(1 - \sum_{j=1}^\xi p_{i,\Phi}[j] \right) \mathbb{P}\big[(g_1^{(i)}, [\mathfrak{o}(g_2^{(i)})]) \big] \\
&\geq \sum_{\mathcal{P}_{\mathbf{g}}} \left(1 - \sum_{j=1}^\xi p_{i,\Phi^*}[j] \right) \mathbb{P}\big[(g_1^{(i)}, [\mathfrak{o}(g_2^{(i)})]) \big] \\
&= L_\xi(\Phi^*, v^*).
\end{aligned}
$$

As Φ, $\mathbf{g}$, and $\mathfrak{o}$ were arbitrary, the proof follows. ∎

Proof [Proof of Remark 18] Fix i, and let $\xi_i = |\Psi(g_1^{(i)}, g_2^{(i)}, v^*)|$. Note that for each $j \leq \xi_i$, the set of graphs $(g_1, g_2) \in \left(g_1^{(i)}, [\mathfrak{o}(g_2^{(i)})]\right)$ for which $\Psi(g_1, g_2, v^*)[j] = \mathcal{I}(\mathfrak{o}(v^*); \mathfrak{o}(g_2))$ is precisely

$$\left(g_1^{(i)}, [\mathfrak{o}(g_2^{(i)})]\right)_{\Psi(g_1^{(i)}, \mathfrak{o}(g_2^{(i)}), v^*)[j] = \mathfrak{o}(v^*)}.$$

If the tie breaking scheme T satisfies $T[1] = \mathfrak{o}(v^*)$, then the set set of graphs $(g_1, g_2) \in \left(g_1^{(i)}, [\mathfrak{o}(g_2^{(i)})]\right)$ for which $\Phi_T^*(g_1, g_2, v^*)[j] = \mathfrak{o}(v^*)$ is then also

$$\left(g_1^{(i)}, [\mathfrak{o}(g_2^{(i)})]\right)_{\Psi(g_1^{(i)}, \mathfrak{o}(g_2^{(i)}), v^*)[j] = \mathfrak{o}(v^*)}.$$

The proof proceeds as follows. For an arbitrary VN scheme Φ, and for each $i \in [h]$,

$$U_{i,\mathbf{g}}^j := \left\{ (g_1, g_2) \in \left(g_1^{(i)}, [\mathfrak{o}(g_2^{(i)})]\right) \text{ s.t. } \mathrm{rank}_{\Phi(g_1, \mathfrak{o}(g_2), v^*)}(\mathfrak{o}(v^*)) = j \right\}$$

$$\subset \left\{ (g_1, g_2) \in \left(g_1^{(i)}, [\mathfrak{o}(g_2^{(i)})]\right) \text{ s.t. } \exists \text{ iso. } \sigma \text{ s.t. } \sigma(\mathfrak{o}(g_2^{(i)})) = \mathfrak{o}(g_2) \text{ and} \right.$$

$$\left. \sigma\left(\mathcal{I}\left(\Phi(g_1^{(i)}, \mathfrak{o}(g_2^{(i)}), v^*)[j]; \mathfrak{o}(g_2^{(i)})\right)\right) \ni \mathfrak{o}(v^*) \right\}$$

$$= \left(g_1^{(i)}, [\mathfrak{o}(g_2^{(i)})]\right)_{\mathcal{I}\left(\Phi(g_1^{(i)}, \mathfrak{o}(g_2^{(i)}), v^*)[j]; \mathfrak{o}(g_2^{(i)})\right) = \mathfrak{o}(v^*)}$$

We then have that $T[1] = \mathfrak{o}(v^*)$ implies that

$$1 - L_\ell(\Phi, v^*) = \sum_{\mathcal{P}_\mathbf{g}} \mathbb{P}\left[\left(\bigcup_{j=1}^\ell U_{i,\mathbf{g}}^j\right) \mid \left(g_1^{(i)}, [\mathfrak{o}(g_2^{(i)})]\right)\right] \mathbb{P}\left[\left(g_1^{(i)}, [\mathfrak{o}(g_2^{(i)})]\right)\right]$$

$$\leq \sum_{\mathcal{P}_\mathbf{g}} \mathbb{P}\left[\bigcup_{j=1}^\ell \left(g_1^{(i)}, [\mathfrak{o}(g_2^{(i)})]\right)_{\mathcal{I}\left(\Phi(g_1^{(i)}, \mathfrak{o}(g_2^{(i)}), v^*)[j]; \mathfrak{o}(g_2^{(i)})\right) = o(v^*)} \mid \left(g_1^{(i)}, [\mathfrak{o}(g_2^{(i)})]\right)\right] \mathbb{P}\left[\left(g_1^{(i)}, [\mathfrak{o}(g_2^{(i)})]\right)\right]$$

$$= \sum_{\mathcal{P}_\mathbf{g}} \sum_{j \in \mathcal{J}_i} \mathbb{P}\left[\left(g_1^{(i)}, [\mathfrak{o}(g_2^{(i)})]\right)_{\mathcal{I}\left(\Phi(g_1^{(i)}, \mathfrak{o}(g_2^{(i)}), v^*)[j]; \mathfrak{o}(g_2^{(i)})\right) = o(v^*)} \mid \left(g_1^{(i)}, [\mathfrak{o}(g_2^{(i)})]\right)\right] \mathbb{P}\left[\left(g_1^{(i)}, [\mathfrak{o}(g_2^{(i)})]\right)\right]$$

$$\leq \sum_{\mathcal{P}_\mathbf{g}} \sum_{j \in \mathcal{J}_i} \underbrace{\mathbb{P}\left[\left(g_1^{(i)}, [\mathfrak{o}(g_2^{(i)})]\right)_{\Psi(g_1^{(i)}, \mathfrak{o}(g_2^{(i)}), v^*)[j] = \mathfrak{o}(v^*)} \mid \left(g_1^{(i)}, [\mathfrak{o}(g_2^{(i)})]\right)\right]}_{=\mathbb{P}\left[(g_1, g_2) \in \left(g_1^{(i)}, [\mathfrak{o}(g_2^{(i)})]\right) \text{ s.t. } \Phi_T^*(g_1, \mathfrak{o}(g_2), v^*)[j] = \mathfrak{o}(v^*) \mid \left(g_1^{(i)}, [\mathfrak{o}(g_2^{(i)})]\right)\right]} \mathbb{P}\left[\left(g_1^{(i)}, [\mathfrak{o}(g_2^{(i)})]\right)\right]$$

$$\leq 1 - L_\ell(\Phi_T^*, v^*),$$

where $\mathcal{J}_i \subset [\ell]$ is the lexicographically smallest set of indices for which

$$\left\{\left(g_1^{(i)}, \mathfrak{o}(g_2^{(i)})]\right)_{\mathcal{I}\left(\Phi(g_1^{(i)}, \mathfrak{o}(g_2^{(i)}), v^*)[j]; \mathfrak{o}(g_2^{(i)})\right) = o(v^*)}\right\}_{j \in \mathcal{J}_i}$$

are distinct. ∎

A.2. Proof of Theorem 23

Proof Define a probability vector $\xi \in \mathbb{R}^m$ by $\xi_i = \epsilon_i - \epsilon_{i-1}$ for $i \in [m-1]$ (where we take $\epsilon_0 := 0$), and let $\xi_m = 1 - \epsilon_{m-1}$. Consider asymmetric graphs $(g_1, g_2) \in \mathcal{G}_n \times \mathcal{G}_m$ and construct a distribution $F_{c,\theta}^{(n,m)} \in \mathcal{N}$ as follows.

i. $c = n \wedge m$;

ii. The support of $F_{c,\theta}^{(n,m)}$ is $(g_1, [\mathfrak{o}(g_2)])$;

iii. For each $k \in [m]$ define

$$R_{\Phi,k} = \big\{ (g_1, \tilde{g}_2) \in (g_1, [\mathfrak{o}(g_2)]) \text{ s.t. } \Phi(g_1, \mathfrak{o}(\tilde{g}_2), v^*)[k] = \mathfrak{o}(v^*) \big\}.$$

Then we define $\mathbb{P}_{F_{c,\theta}^{(n,m)}}(R_{\Phi,k}) := \xi[k]$ with all elements of $R_{\Phi,k}$ being assigned equal mass under $F_{c,\theta}^{(n,m)}$.

It is clear then that $L_k(\Phi, v^*) = 1 - \epsilon_k > 1 - \frac{k}{m}$. It is also clear that $L_k^*(v^*) \leq \epsilon_{m-k}$. Indeed, consider Φ' which is defined by reversing the order provided by Φ; then $L_k(\Phi', v^*) = \epsilon_{m-k}$; which completes the proof. $\blacksquare$

Appendix B. Proof of Theorem 19

Herein we will provide a sketch of the proof of Theorem 19 for completeness. Let Q be a permutation matrix in Π_n (with associated permutation σ_Q) that permutes precisely k labels (i.e., $\sum_i Q_{i,i} = n - k$), and let T denote the number of transpositions induced by σ_Q, and note $|T| \leq k/2$. By exploiting the cyclic structure of Q acting on $\text{vec}(B)$, we have that

$$\mathbb{E}\delta(Q) - \mathbb{E}\delta(I_n) = \mathbb{E}\|AQ - QB\|_F^2 - \mathbb{E}\|A - B\|_F^2 \geq \epsilon \left((n-k)k + \binom{k}{2} - T \right).$$

Combining this expectation bound with the following McDiarmid-like concentration result will yield the proof of Theorem 19.

Proposition 29 (Proposition 3.2 from Kim et al. (2002)) *Let $X_1, \ldots, X_n$ be a sequence of independent Bernoulli random variables where $\mathbb{E}[X_i] = p_i$. Let $f : \{0,1\}^n \mapsto \Re$ be such that changing any X_i to $1 - X_i$ changes f by at most*

$$M = \sup_i \sup_{X_1, \ldots, X_n} |f(X_1, \ldots, X_i, \ldots, X_n) - f(X_1, \ldots, 1 - X_i, \ldots, X_n)|.$$

Let $\sigma^2 = M^2 \sum_i p_i(1 - p_i)$ and let $Y = f(X_1, \ldots, X_n)$.
   Then

$$\Pr[|Y - \mathbb{E}[Y]| \geq t\sigma] \leq 2e^{-t^2/4}$$

for all $0 < t < 2\sigma/M$.

Indeed, we see that $X_Q := \delta(Q) - \delta(I_n)$ is a function of N_Q independent Bernoulli random variables, where $N_Q = 3\left(\binom{k}{2} + k(n-k) - |T|\right) \leq 3kn$. Let S_Q be the sum of these N_Q Bernoulli random variables, and it follows that $\text{Var}(S_Q) \leq N_Q/4$. By setting $t = C\frac{\epsilon kn}{\sigma}$ for an appropriate constant $C > 0$ in Proposition 29, we have for n sufficiently large

$$\Pr\left(X_Q \leq 0\right) \leq \Pr\left(|X_Q - \mathbb{E}(X_Q)| \geq \mathbb{E}(X_Q)\right) \leq 2\exp\left\{-\Theta(\epsilon^2 kn)\right\}.$$

A union bound over all such Q (of which there are $\leq n^k$) and over k yields

$$\mathbb{P}\left(\exists\, Q \in \Pi_n \setminus \{I_n\} \text{ s.t. } \delta(Q) \leq \delta(I_n)\right) \leq \sum_{k=2}^{n} 2\exp\left\{k\log(n) - \Theta(\epsilon^2 kn)\right\} = 2\exp\left\{-\Theta(\epsilon^2 n)\right\},$$

as desired.

B.1. VN Schemes with Ties

We can incorporate ties into the VN framework as follows. With ties allowed, any sensibly-defined vertex nomination scheme should view all vertices in $\mathcal{I}(u; g)$ as being equally good matches to a vertex of interest v^*. To this end, we will view VN schemes as providing *weak orderings* of the elements of W:

Definition 30 *For a set A, let $\mathfrak{W}_A$ denote the set of all weak orderings of the elements of A (i.e., the set of all total orderings where ties are allowed). For $x \in \mathfrak{W}_A$, let t_x be any maximum-length total ordering induced by x. For each $a \in A$, we define*

$$rank_x(a) = rank_{t_x}(a'),$$

where $a = a'$ according to the ordering x.

Example 7 *If $A = \{a, b, c, d, e\}$ and $x : a > c > d = e > b$, then $t(x) : a > c > d > b$, or $t(x) : a > c > e > b$; in either case, $rank_x(a) = 1$, $rank_x(c) = 2$, $rank_x(d) = 3$, $rank_x(e) = 3$, and $rank_x(b) = 4$.*

A well-defined VN scheme should be "label-independent" in the following sense: Each element of each $\mathcal{I}(\mathfrak{o}(u); \mathfrak{o}(g_2))$ should be ranked identically by Φ, and these ranks should be independent of the obfuscation function $\mathfrak{o}$. Formally, we have the following.

Definition 31 (Vertex Nomination (VN) Scheme) *For $n, m > 0$ fixed, obfuscating set W, and $\mathfrak{o} \in \mathfrak{D}_W$, a vertex nomination scheme is a function $\Phi : \mathcal{G}_n \times \mathfrak{o}(\mathcal{G}_m) \times V_1 \to \mathfrak{W}_W$ satisfying the following properties: For all $(g_1, g_2) \in \mathcal{G}_n \times \mathcal{G}_m$,*

    i. If $u_1 \notin \mathcal{I}(u_2; g_2)$ then either $\mathfrak{o}(u_1) > \mathfrak{o}(u_2)$ or $\mathfrak{o}(u_1) < \mathfrak{o}(u_2)$ in the ordering provided by $\Phi(g_1, \mathfrak{o}(g_2), v^*)$;

    ii. If $u_1 \in \mathcal{I}(u_2; g_2)$ then $\mathfrak{o}(u_1) = \mathfrak{o}(u_2)$ in the ordering provided by $\Phi(g_1, \mathfrak{o}(g_2), v^*)$;

    iii. (consistency criterion) If $\mathfrak{o}_1, \mathfrak{o}_2 \in \mathfrak{D}_W$, then for each $v \in V(g_2)$

$$rank_{\Phi(g_1, \mathfrak{o}_1(g_2), v^*)}(\mathfrak{o}_1(v)) = rank_{\Phi(g_1, \mathfrak{o}_2(g_2), v^*)}(\mathfrak{o}_2(v)). \tag{14}$$

We let $\mathcal{V}_{n,m}$ denote the set of all such VN schemes.

The VN loss functions and level-k errors are defined as in the totally ordered setting. To define the Bayes optimal scheme, let (g_1, g_2) be realized from $(G_1, G_2) \sim F_{c,\theta}^{(n,m)} \in \mathcal{N}_{n,m}$, and consider a vertex of interest $v^* \in C$ and obfuscating function $\mathfrak{o} : V_2 \to W$. Letting $\simeq$ denote graph isomorphism, for $(g_1, g_2) \in \mathcal{G}_n \times \mathcal{G}_m$ define the set

$$
\begin{aligned}
(g_1, [\mathfrak{o}(g_2)]) &= \left\{ (g_1, \tilde{g}_2) \text{ s.t. } \mathfrak{o}(\tilde{g}_2) \simeq \mathfrak{o}(g_2) \right\} \\
&= \left\{ (g_1, \tilde{g}_2) \text{ s.t. } \tilde{g}_2 \simeq g_2 \right\}
\end{aligned}
\tag{15}
$$

In order to define the Bayes optimal scheme, we will also need the following restrictions of $(g_1, [\mathfrak{o}(g_2)])$: for each $w \in W$, and $v \in V_2$ we define

$$
\begin{aligned}
(g_1, [\mathfrak{o}(g_2)])_{\mathcal{I}(w;\mathfrak{o}(g_2))=\mathfrak{o}(v)} = \Big\{ &(g_1, \tilde{g}_2) \text{ s.t. } \exists \text{ iso. } \sigma \text{ with } \mathfrak{o}(\tilde{g}_2) = \sigma(\mathfrak{o}(g_2)), \\
&\text{and } \sigma(u) = \mathfrak{o}(v) \text{ for some } u \in \mathcal{I}(w; \mathfrak{o}(g_2)) \Big\}.
\end{aligned}
\tag{16}
$$

Note that for a fixed v, if $\{\mathcal{I}(w; \mathfrak{o}(g_2))\}_{w \in W'}$ partitions W (for some suitable $W' \subseteq W$), then

$$
\left\{ (g_1, [\mathfrak{o}(g_2)])_{\mathcal{I}(w;\mathfrak{o}(g_2))=\mathfrak{o}(v)} \right\}_{w \in W'}
$$

partitions $(g_1, [\mathfrak{o}(g_2)])$. We are now ready to define a Bayes optimal VN scheme.

For ease of notation, in the sequel we will write $\mathbb{P}_{F_{c,\theta}^{(n,m)}}$ or even simply $\mathbb{P}$ in place of $\mathbb{P}_{(G_1,G_2) \sim F_{c,\theta}^{(n,m)}}$ where there is no risk of ambiguity. Let

$$
\mathbf{g} = \left\{ \left(g_1^{(i)}, g_2^{(i)} \right) \right\}_{i=1}^{h}
\tag{17}
$$

be such that the sets

$$
\left\{ \left(g_1^{(i)}, [\mathfrak{o}(g_2^{(i)})] \right) \right\}_{i=1}^{h}
$$

partition $\mathcal{G}_n \times \mathcal{G}_m$. We will call this partition $\mathcal{P}_{n,m}$, where we suppress dependence on $\mathbf{g}$ and $\mathfrak{o}$ for ease of notation. We will define a Bayes optimal scheme $\Phi^* = \Phi_{\mathbf{g}}^*$ piecewise on each element of this partition. For each $i \in [h]$, define (where ties in the argmax's are broken in an arbitrary but nonrandom manner)

$$
\Phi^*\left(g_1^{(i)}, \mathfrak{o}(g_2^{(i)}), v^* \right)[1] \in \operatorname*{argmax}_{\mathcal{I}(u;\mathfrak{o}(g_2^{(i)})) \subset W} \mathbb{P}\left[\left(g_1^{(i)}, [\mathfrak{o}(g_2^{(i)})] \right)_{\mathcal{I}(u;\mathfrak{o}(g_2^{(i)}))=\mathfrak{o}(v^*)} \, \Bigg| \, \left(g_1^{(i)}, [\mathfrak{o}(g_2^{(i)})] \right) \right]
$$

$$
\Phi^*\left(g_1^{(i)}, \mathfrak{o}(g_2^{(i)}), v^* \right)[2] \in \operatorname*{argmax}_{\mathcal{I}(u;\mathfrak{o}(g_2^{(i)})) \subset W \setminus \{\Phi^*[1]\}} \mathbb{P}\left[\left(g_1^{(i)}, [\mathfrak{o}(g_2^{(i)})] \right)_{\mathcal{I}(u;\mathfrak{o}(g_2^{(i)}))=\mathfrak{o}(v^*)} \, \Bigg| \, \left(g_1^{(i)}, [\mathfrak{o}(g_2^{(i)})] \right) \right]
$$

$$
\vdots
$$

$$
\Phi^*\left(g_1^{(i)}, \mathfrak{o}(g_2^{(i)}), v^* \right)[\ell] \in \operatorname*{argmax}_{\mathcal{I}_{\mathfrak{o}(g_2)}(u) \subset W \setminus \{\cup_{j=1}^{\ell-1} \Phi^*[j]\}} \mathbb{P}\left[\left(g_1^{(i)}, [\mathfrak{o}(g_2^{(i)})] \right)_{\mathcal{I}(u;\mathfrak{o}(g_2^{(i)}))=\mathfrak{o}(v^*)} \, \Bigg| \, \left(g_1^{(i)}, [\mathfrak{o}(g_2^{(i)})] \right) \right]
$$

$$
\tag{18}
$$

so that the ranking provided by Φ^* is (where $\Phi^*(g_1^{(i)}, \mathfrak{o}(g_2^{(i)}), v^*)[i] = \{u_1^{(i)}, \ldots, u_{n_i}^{(i)}\}$)

$$\underbrace{u_1^{(1)} = u_2^{(1)} = \cdots = u_{n_1}^{(1)}}_{n_1} > \underbrace{u_1^{(2)} = u_2^{(2)} = \cdots = u_{n_2}^{(2)}}_{n_2} > \cdots > \underbrace{u_1^{(\ell)} = u_2^{(\ell)} = \cdots = u_{n_\ell}^{(\ell)}}_{n_\ell}.$$

For each element

$$(g_1', g_2') \in (g_1^{(i)}, [\mathfrak{o}(\tilde{g}_2^{(i)})]) \setminus \{(g_1^{(i)}, g_2^{(i)})\},$$

choose any isomorphism σ such that $\mathfrak{o}(g_2') = \sigma(\mathfrak{o}(g_2^{(i)}))$, and define

$$\Phi^*(g_1, \mathfrak{o}(g_2'), v^*) = \sigma(\Phi^*(g_1^{(i)}, \mathfrak{o}(g_2^{(i)}), v^*)),$$

noting that $\Phi^*(g_1, \mathfrak{o}(g_2'), v^*)$ is independent of the choice of isomorphism σ. The next proposition states that, modulo ties, the definition of $\Phi_{\mathbf{g}}^*$ is independent of the choice of $\mathbf{g}$.

Proposition 32 *Let $\mathfrak{o} \in \mathfrak{O}_W$ be an obfuscating function, and let*

$$\mathbf{g} = \left\{ \left(g_1^{(i)}, g_2^{(i)} \right) \right\}_{i=1}^h \neq \tilde{\mathbf{g}} = \left\{ \left(g_1^{(i)}, \tilde{g}_2^{(i)} \right) \right\}_{i=1}^h$$

be such that the sets

$$\left\{ \left(g_1^{(i)}, [\mathfrak{o}(g_2^{(i)})] \right) \right\}_{i=1}^h, \left\{ \left(g_1^{(i)}, [\mathfrak{o}(\tilde{g}_2^{(i)})] \right) \right\}_{i=1}^h$$

partition $\mathcal{G}_n \times \mathcal{G}_m$. Suppose that $(G_1, G_2) \sim F_{c,\theta}^{(n,m)} \in \mathcal{N}_{n,m}$, and consider a vertex of interest $v^* \in C$. Then there exists a fixed strategy for breaking ties in the argmax's for $\Phi_{\mathbf{g}}^*$ and $\Phi_{\tilde{\mathbf{g}}}^*$ that yields $\Phi_{\mathbf{g}}^* = \Phi_{\tilde{\mathbf{g}}}^*$. In particular, under any such tie-breaking strategy, we have that $L_k(\Phi_{\mathbf{g}}^*, v^*) = L_k(\Phi_{\tilde{\mathbf{g}}}^*, v^*)$ for all $k \in [m-1]$.

Lastly, the following theorem shows that this scheme (or schemes) is indeed Bayes optimal. The proof is analogous to the totally ordered setting and is thus omitted.

Theorem 33 *Let $\mathfrak{o} \in \mathfrak{O}_W$ be an obfuscating function, and let*

$$\mathbf{g} = \left\{ \left(g_1^{(i)}, g_2^{(i)} \right) \right\}_{i=1}^h$$

be such that the sets

$$\left\{ \left(g_1^{(i)}, [\mathfrak{o}(g_2^{(i)})] \right) \right\}_{i=1}^h$$

partition $\mathcal{G}_n \times \mathcal{G}_m$. Let $\Phi^* = \Phi_{\mathbf{g}}^*$ be as defined in Equation (9). Suppose that $(G_1, G_2) \sim F_{c,\theta}^{(n,m)} \in \mathcal{N}_{n,m}$, and consider a vertex of interest $v^* \in C$. We have that $L_k(\Phi^*, v^*) = L_k^*(v^*)$ for all $k \in [m-1]$ and all obfuscating functions $\mathfrak{o}$.

References

A. Agarwal, S. Chakrabarti, and S. Aggarwal. Learning to rank networked entities. In *Proceedings of the ACM Conference on Knowledge Discovery and Data Mining*, pages 14–23, 2006.

S. Agarwal. Learning to rank on graphs. *Machine Learning*, 10(3):333–357, 2010.

E. M. Airoldi, D. M. Blei, S. E. Fienberg, and E. P. Xing. Mixed membership stochastic blockmodels. *The Journal of Machine Learning Research*, 9:1981–2014, 2008.

R. Albert and A.-L. Barabási. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74(1):47, 2002.

D. Asta and C. R. Shalizi. Geometric network comparison. In M. Meila and T. Haskes, editors, *Proceedings of the 31st Conference on Uncertainty in Artificial Intelligence*, pages 102–110, 2015.

M. Belkin and P. Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15(6):1373–1396, 2003.

M. Belkin, P. Niyogi, and V. Sindhwani. Manifold Regularization: A Geometric Framework for Learning from Examples. *Journal of Machine Learning Research*, 7:2399–2434, 2006.

P. J. Bickel and A. Chen. A nonparametric view of network models and Newman-Girvan and other modularities. *Proc. National Academy of Sciences, USA*, 106:21068–21073, 2009.

S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine. In *Proceedings of the 7th International World Wide Web Conference*, 1998.

L. Chen, C. Shen, J. T. Vogelstein, and C. E. Priebe. Robust vertex classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(3):578–590, 2016a.

L. Chen, J. T. Vogelstein, V. Lyzinski, and C. E. Priebe. A joint graph inference case study: The C. elegans chemical and electrical connectomes. *Worm*, 5(2), 2016b.

R. R. Coifman and S. Lafon. Diffusion maps. *Applied and Computational Harmonic Analysis*, 21:5–30, 2006.

D. Conte, P. Foggia, C. Sansone, and M. Vento. Thirty years of graph matching in pattern recognition. *International Journal of Pattern Recognition and Artificial Intelligence*, 18 (03):265–298, 2004.

G. Coppersmith. Vertex nomination. *Wiley Interdisciplinary Reviews: Computational Statistics*, 6(2):144–153, 2014.

G. A. Coppersmith and C. E. Priebe. Vertex nomination via content and context. *arXiv preprint arXiv:1201.4118*, 2012.

L. Devroye, L. Györfi, and G. Lugosi. *A Probabilistic Theory of Pattern Recognition.* Springer, 1997.

K. K. Duh. *Learning to Rank with Partially-Labeled Data.* PhD thesis, University of Washington, 2009.

P. Erdős and A. Rényi. On random graphs, I. *Publicationes Mathematicae*, 6:290–297, 1959.

P. Erdős and A. Rényi. Asymmetric graphs. *Acta Mathematica Academiae Scientiarum Hungarica*, 14(3–4):295–315, 1963.

D. E. Fishkind, V. Lyzinski, H. Pao, L. Chen, and C. E. Priebe. Vertex nomination schemes for membership prediction. *The Annals of Applied Statistics*, 9(3):1510–1532, 2015.

D. E. Fishkind, S. Adali, H. G. Patsolic, L. Meng, D. Singh, V. Lyzinski, and C. E. Priebe. Seeded graph matching. *Pattern Recognition*, 87:203–215, 2019.

O. Frank and D. Strauss. Markov graphs. *Journal of the american Statistical association*, 81(395):832–842, 1986.

P. D. Hoff, A. E. Raftery, and M. S. Handcock. Latent space approaches to social network analysis. *Journal of the American Statistical Association*, 97(460):1090–1098, 2002.

P. W. Holland, K. B. Laskey, and S. Leinhardt. Stochastic blockmodels: First steps. *Social Networks*, 5(2):109–137, 1983.

H. Jeong, S. P. Mason, A.-L. Barabási, and Z. N. Oltvai. Lethality and centrality in protein networks. *Nature*, 411(6833):41–42, 2001.

B. Karrer and M. E. J. Newman. Stochastic blockmodels and community structure in networks. *Physical Review E*, 83, 2011.

J. H. Kim, B. Sudakov, and V. H. Vu. On the asymmetry of random regular graphs and random graphs. *Random Structures and Algorithms*, 21:216–224, 2002.

H. Li. Learning to rank for information retrieval and natural language processing. *Synthesis Lectures on Human Language Technologies*, 4(1), 2011.

D. Liben-Nowell and J. Kleinberg. The link-prediction problem for social networks. *Journal of the Association for Information Science and Technology*, 58(7):1019–1031, 2007.

T.-Y. Liu. Learning to rank for information retrieval. *Foundations and Trends in Information Retrieval*, 3(3):225–331, 2009.

L. Lü and T. Zhou. Link prediction in complex networks: A survey. *Physica A: Statistical Mechanics and its Applications*, 390(6):1150–1170, 2011.

C. Luo, W. Pang, Z. Wang, and C. Lin. Hete-CF: Social-based collaborative filtering recommendation using heterogeneous relations. In *Proceedings of the IEEE International Conference on Data Mining*, pages 917–922, 2014.

D. Lusseau and M. E. J. Newman. Identifying the role that animals play in their social networks. *Proceedings of the Royal Society of London B: Biological Sciences*, 271(Suppl 6):S477–S481, 2004.

U. Von Luxburg. A tutorial on spectral clustering. *Statistics and computing*, 17(4):395–416, 2007.

V. Lyzinski. Information recovery in shuffled graphs via graph matching. *IEEE Transactions on Information Theory*, 64(5):3254–3273, 2018.

V. Lyzinski and D. L. Sussman. Matchability of heterogeneous networks pairs. *arXiv preprint arXiv:1705.02294*, 2018.

V. Lyzinski, D. L. Sussman, M. Tang, A. Athreya, and C. E. Priebe. Perfect clustering for stochastic blockmodel graphs via adjacency spectral embedding. *Electronic Journal of Statistics*, 8:2905–2922, 2014.

V. Lyzinski, D. Fishkind, M. Fiori, J.T. Vogelstein, C.E. Priebe, and G. Sapiro. Graph matching: Relax at your own risk. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, In press, 2015.

V. Lyzinski, K. Levin, D. E. Fishkind, and C. E. Priebe. On the consistency of the likelihood maximization vertex nomination scheme: Bridging the gap between maximum likelihood estimation and graph matching. *Journal of Machine Learning Research*, 17(179):1–34, 2016.

H. Ma, I. King, and M. R. Lyu. Mining web graphs for recommendations. *IEEE Transactions on Knowledge and Data Engineering*, 24(6):1051–1064, 2012.

C. Manning, P. Raghavan, and H. Schütze. *Introduction to Information Retrieval*. Cambridge University Press, 2008.

D. Marchette, C. E. Priebe, and G. Coppersmith. Vertex nomination via attributed random dot product graphs. In *Proceedings of the 57th ISI World Statistics Congress*, volume 6, page 16, 2011.

R. Mihalcea and D. Radev. *Graph-based Natural Language Processing and Information Retrieval*. Cambridge University Press, 2011.

M. E. J. Newman. A measure of betweenness centrality based on random walks. *Social Networks*, 27(1):39–54, 2005.

M. E. J. Newman. Finding community structure in networks using the eigenvectors of matrices. *Phys. Rev. E*, 74(3):036104, 2006.

M. E. J. Newman and A. Clauset. Structure and inference in annotated networks. *Nature Communications*, 7(11863), 2016.

H. G. Patsolic, Y. Park, V. Lyzinski, and C. E. Priebe. Vertex nomination via local neighborhood matching. *arXiv preprint arXiv:1705.00674*, 2017.

T. A. N. Pham, X. Li, G. Cong, and Z. Zhang. A general recommendation model for heterogeneous networks. *IEEE Transactions on Knowledge and Data Engineering*, 28 (12):3140–3153, 2016.

P. Resnick and H. R. Varian. Recommender systems. *Communications of the ACM*, 40(3): 56–58, 1997.

F. Ricci, L. Rokach, and B. Shapira. *Introduction to Recommender Systems Handbook*. Springer, 2011.

G. Robins, P. Pattison, Y. Kalish, and D. Lusher. An introduction to exponential random graph (p^*) models for social networks. *Social Networks*, 29(2):173–191, 2007.

K. Rohe, S. Chatterjee, and B. Yu. Spectral clustering and the high-dimensional stochastic blockmodel. *Annals of Statistics*, 39:1878–1915, 2011.

S. Rothe and H. Schütze. CoSimRank: A flexible & efficient graph-theoretic similarity measure. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, pages 1392–1402, 2014.

T. Snijders, P. Pattison, G. Robins, and M. Handcock. New specifications for exponential random graph models. *Sociological Methodology*, 36(1):99–153, 2006.

I. Steinwart. Support vector machines are universally consistent. *Journal of Complexity*, 18(3):768–791, 2002.

C. Stone. Consistent nonparametric regression. *Annals of Statistics*, 5:595–645, 1977.

M. Sun and C. E. Priebe. Efficiency investigation of manifold matching for text document classification. *Pattern Recognition Letters*, 34(11):1263–1269, 2013.

D. L. Sussman, M. Tang, D. E. Fishkind, and C. E. Priebe. A consistent adjacency spectral embedding for stochastic blockmodel graphs. *Journal of the American Statistical Association*, 107(499):1119–1128, 2012.

D. L. Sussman, V. Lyzinski, Y. Park, and C. E. Priebe. Matched filters for noisy induced subgraph detection. *arXiv preprint arXiv:1803.02423*, 2018.

S. Suwan, D. S. Lee, and C. E. Priebe. Bayesian vertex nomination using content and context. *Wiley Interdisciplinary Reviews: Computational Statistics*, 7(6):400–416, 2015.

M. Tang, D. L. Sussman, and C. E. Priebe. Universally consistent vertex classification for latent positions graphs. *The Annals of Statistics*, 41(3):1406–1430, 2013.

M. Tang, A. Athreya, D. L. Sussman, V. Lyzinski, Y. Park, and C. E. Priebe. A semiparametric two-sample hypothesis testing problem for random graphs. *Journal of Computational and Graphical Statistics*, 26(2):344–354, 2017a.

M. Tang, A. Athreya, D. L. Sussman, V. Lyzinski, and C. E. Priebe. A nonparametric two-sample hypothesis testing problem for random dot product graphs. *Bernoulli*, 23(3): 1599–1630, 2017b.

D. Watts and S. H. Strogatz. Collective dynamics of small-world networks. *Nature*, 393 (6684):440, 1998.

J. Yoder, L. Chen, H. Pao, E. Bridgeford, K. Levin, D. E. Fishkind, C. E. Priebe, and V. Lyzinski. Vertex nomination: The canonical sampling and the extended spectral nomination schemes. *arXiv preprint arXiv:1802.04960*, 2018.

S. Young and E. Scheinerman. Random dot product graph models for social networks. In *Proceedings of the 5th International Conference on Algorithms and Models for the Web-graph*, pages 138–149, 2007.

D. Zhou, O. Bousquet, T. N. Lal, J. Weston, and B. Schölkopf. Learning with local and global consistency. In S. Thrun, L. K. Saul, and P. B. Schölkopf, editors, *Advances in Neural Information Processing Systems 16*, pages 321–328, 2004.