Ranking Networks

Amirali Salehi-Abari
Department of Computer Science
University of Toronto
abari@cs.toronto.edu

Craig Boutilier
Department of Computer Science
University of Toronto
cebly@cs.toronto.edu

Abstract

Latent space models for network formation assume that nodes possess latent attributes that determine their propensity to connect. We propose a new model for network formation, *ranking networks*, in which these attributes are *rankings* over some space of alternatives. Such rankings may reflect user preferences, relevance/quality judgements, etc., while ranking networks capture correlations of, say, user preferences across a social network. We present preliminary theoretical and empirical analyses of structural properties of such networks, and develop algorithmic approximations to help efficiently predict these properties. Empirical results demonstrate the quality of these approximations.

1 Introduction

Network formalisms offer a powerful tool for studying the interactions of entities in complex systems. Of special interest in network science is the development of *network formation models* that explain the emergence of common structural properties of real-world networks [2, 34]. Recent work has focused on modeling the role of both known and hidden attributes of nodes in forming networks—the intuition being that each node possesses attributes (e.g., geographical location [35] or social position [14, 7]) that determine the chance that any pair of nodes become connected. Examples of such models include spatial networks [4], random geometric graphs [31], latent space models [14, 7], and multiplicative attribute graphs [19].

Most of these models take attributes to be binary, real-valued, or integral. Surprisingly, little work has studied attribute-based network formation in which the underlying attributes correspond to *rankings* over some space of options. Such rankings may represent the preferences of agents over some set of alternatives (e.g., products, political candidates, jobs) or their subjective opinions about the quality or relevance of certain items (web pages, sports teams). Because of the increasing availability of ranked data due to recommender systems, web search, online surveying, and other applications, a deeper understanding of the possible interactions between ranked data and network formation seems imperative. In this work, we develop and begin a preliminary analysis of *ranking networks*, in which each node in the network possesses a ranking, and the similarity between two nodes' rankings determines their affinity or probability of connection. Such networks can be used for a variety of purposes, but we are especially interested in models where rankings reflect *agent preferences* over a some space of options. We discuss the extent to which we can leverage the structure of ranking networks to address algorithmic questions that arise in the analysis of networks. For instance, what topological network properties emerge under such a model? How to model parameters (including the ranking distribution and similarity metric) influence these properties? And can such structural properties be predicted efficiently?

Ranking network models can be deployed in a variety of contexts. It is widely acknowledged that the behaviours of individuals are correlated over the social networks [10, 15]; and correlated preferences offer one possible explanation of this phenomenon. Ranking networks can help predict or elicit the preferences of specific individuals to develop more efficient recommender, social choice and

advertising mechanisms. They may also have application in information retrieval and topic modeling over the web: given some common set of attributes (e.g., topics, tags), each node (e.g., web page, photo) may have a "relevance ranking" over those attributes. A ranking network can capture the correlation of such relevance rankings given the structure of underlying information network (and may help improve the efficiency of recently developed comparison-based interactive search methods [16]).The predictive power of ranking networks might be also harnessed for link prediction [24, 22]—to help predict future interactions between entities—or *link mining*—to help infer unobserved links between entities [13].

We introduce a probabilistic model for the formation of *ranking networks*, in which each entity possesses a ranking over a set of alternatives. The similarity of their rankings determines the chance of connectivity between two individuals. We first analyze some general topological properties of ranking networks (e.g., connectivity conditions, graph diameter, degree distribution, edge density, clustering coefficient). We then study distance-based ranking networks in which similarity is measured using standard ranking distance metrics, examine a few additional properties in this setting, and describe several algorithmic approximations to efficiently predict structural properties of the resulting network. Our approximation methods are also applicable for learning model parameters and inference problem. We present preliminary empirical results using the *Mallows' ϕ-model* [27], a distributional model of rankings used widely in psychometrics and machine learning, and well-suited to modeling user preferences. These results confirm the efficiency and predictive accuracy of our computational approximations.

2 Related Work

The pioneering random graph model, $G(n, p)$, of Erdős and Rènyi [11] assumes an edge between any pair of nodes occurs with probability p. Other seminal work includes the preferential attachment [2] and small-world [34] models. Since then network formation models have drawn considerable attention, with a variety of models proposed to capture structural properties of real-world networks [29, 3, 30].One can roughly divide these model into two categories. *Static models* (e.g., Erdős-Rènyi, small world) specify the network using simple static properties (e.g., probability of edge occurrence), whereas *dynamic models* (e.g., preferential attachment) involve some dynamic "growth" process by which nodes and edges are added to a network.

Our ranking network model is a random, static model. It is also falls within the class of *spatial (or latent space) networks* [4, 14] in which nodes have a set of real-valued, binary, or integer-valued *latent variables*, with the probability of an edge forming between two nodes determined by their attributes. The *Waxman random graph model* [35], often used to model the computer networks, distributes n nodes uniformly at random on the plane, with two nodes connected with probability that decreases exponentially with their Euclidean distance. Hoff *et al.* [14] develop a similar model for inferring latent variables on social networks—nodes are points in a d-dimensional Euclidean "social space." The *random dot product model* [36] is somewhat distinct in that the distance between nodes is given by the inner product of their position vectors. The *hidden variable model* [6, 7] is a generalization of the Waxman model: nodes are equipped with a hidden (real-valued or integer) random variable drawn independently from a specified distribution. Two nodes are connected according to a symmetric probability function over node attributes. Serrano et al.[33] slightly modify this approach to introduce a general class of models based on a hidden metric space where nodes—located at a specific point in this space—are connected with a probability determined by a *connection probability function* defined over node distances. The *multiplicative attribute graphs (MAG) model* [19] generalizes the Kronecker graph model [20], assigning each node a vector of categorical attributes, and using a link-affinity matrix to capture connection probabilities based on attribute interactions.

Within spatial network formation models, recent research has addressed *inference and learning* of latent attributes and social network structure. Hoff *et al.* [14] develop inference and learning techniques to infer node locations in a fixed d-dimensional Euclidean "social space." Using the MAG model, Kim and Leskovec [17] develop a scalable *variational expectation maximization* method for learning model parameters given network structure and node attributes, and inferring node's latent variables and model parameters when the network structure is observed. Kim and Leskovec [18] extend this to allow nodes to have multiple latent characteristics. Other work on learning on graphs includes *collective classification* [32] and active learning over network data [5, 26].

3 The Formation of Ranking Networks

We first provide background on latent space models for network formation. We then define *ranking networks* by building upon these models and discuss some general theoretical properties. Finally, we examine the special case of *distance-based ranking networks*, analysing some of its theoretical properties and developing computational approximations to estimate these properties.

3.1 Latent Space Network Models

Latent space or hidden variable network models [6] generally assume a set n nodes, where each node i is associated with a latent variable h_i. This represents some feature (or feature vector) of the node or individual in question. The parameters of such a model are given by a parameter vector $\boldsymbol{\theta}$ (we discuss model-specific parameterizations below). A random undirected graph is generated by assuming: (1) a distribution $\rho(h|\boldsymbol{\eta})$ specifying the probability that an arbitrary node takes on value h; and (2) a symmetric *connection probability function*, $c(h, h^{'})$, where $c(h_i, h_j)$ denotes the probability that edge e_{ij} forms between nodes i and j given their variable values. A common variant of this model defines the connection probability using the relative distance d_{ij} between the values h_i and h_j in some metric space. In this case, we assume a function $c(d) : [0, \infty) \rightarrow [0, 1]$ that maps distances into connection probabilities. We generally assume distance-based connection functions below. A canonical example of such a connection probability is Waxman's [35] function $c(d) = \beta e^{-\frac{d}{d_0}}$, widely used for modeling computer network topologies, where β controls edge density and d_0 represents a "typical" distance between nodes. Another is $c(d) = (1 + \frac{d}{\beta})^{-\alpha}$, used for modeling social networks [7]. Here β controls the average node degree while $\alpha > 1$ determines the degree of homophily. Let $\boldsymbol{\lambda}$ denote the (vector of) model parameters governing connection function $c(\cdot)$ and $\boldsymbol{\theta} = (\boldsymbol{\eta}, \boldsymbol{\lambda})$ the vector of all parameters of a latent space model.

We now describe certain structural properties that emerge under latent space models. Because we will deal with discrete ranking spaces below, we assume latent variables are discrete. First, note that *edge density*, i.e., the probability of an occurring edge between two arbitrary nodes, is given by:

$$\mathcal{E}(\boldsymbol{\theta}) = \sum_{h,h^{'}} \rho(h|\boldsymbol{\eta})\rho(h^{'}|\boldsymbol{\eta})c\left(d(h^{'}, h)|\boldsymbol{\lambda}\right). \tag{1}$$

As a result, the expected number of edges of a random node is given by $\binom{n}{2}\mathcal{E}(\boldsymbol{\theta})$. The probability of edge between a node i with observed value h_i and a random alternative node j is:

$$\mathcal{D}(h_i, \boldsymbol{\theta}) = \sum_{h^{'}} \rho(h^{'}|\boldsymbol{\eta})c\left(d(h^{'}, h_i)|\boldsymbol{\lambda}\right). \tag{2}$$

Thus, the expected degree of a node with value h_i is given by $(n-1)\mathcal{D}(h_i, \boldsymbol{\theta})$. Note that $\mathcal{E}(\boldsymbol{\theta})$ can be rewritten as:

$$\mathcal{E}(\boldsymbol{\theta}) = \sum_{h} \rho(h|\boldsymbol{\eta})\mathcal{D}(h, \boldsymbol{\theta}). \tag{3}$$

The exact *degree distribution* $P(k, \boldsymbol{\theta})$ is:

$$P(k, \boldsymbol{\theta}) = \sum_{h} \rho(h|\boldsymbol{\eta})\mathcal{G}(h, k|\boldsymbol{\theta}), \tag{4}$$

where $\mathcal{G}(h, k|\boldsymbol{\theta})$ is the probability that a node with value h has k neighbours:

$$\mathcal{G}(h, k|\boldsymbol{\theta}) = \binom{n-1}{k}\mathcal{D}(h, \boldsymbol{\theta})^k \left(1 - \mathcal{D}(h, \boldsymbol{\theta})\right)^{n-1-k}. \tag{5}$$

The *clustering coefficient* of a node i is the fraction of the *pairs* of neighboring nodes that are themselves connected. The expected clustering coefficient for a node i with value h_i can be computed by:

$$\mathcal{C}(h_i, \boldsymbol{\theta}) = \frac{1}{\mathcal{D}(h_i, \boldsymbol{\theta})^2} \sum_{h^{'}, h^{''}} \rho(h^{'})\rho(h^{''})c(d(h_i, h^{'}))c(d(h_i, h^{''}))c(d(h^{'}, h^{''})) \tag{6}$$

One can compute the average clustering by $\langle \mathcal{C}(\boldsymbol{\theta}) \rangle = \sum_h \rho(h, \boldsymbol{\eta})\mathcal{C}(h, \boldsymbol{\theta})$.

3.2 Ranking Networks: A General Model

We adapt the general latent space network model to setting where the latent attributes are *rankings* over some set of alternatives. For instance, nodes might represent individuals with attributes reflecting their preferences over some set of products, services, political candidates, etc. We assume a finite set of alternatives (or options) $\mathcal{A} = \{a_1, \ldots, a_m\}$ and a set of nodes $\mathcal{N} = \{1, \ldots, n\}$. Each node i has a ranking (or strict total order) over $\mathcal{A}$, denoted by $\succ_i$ (weaker notions, e.g., preorders, partial orders, can be accommodated, though some details of our model require modification). Let $\Omega(\mathcal{A})$ denote the set of all $m!$ rankings over $\mathcal{A}$. Our latent variables are rankings r drawn from $\Omega(\mathcal{A})$. We assume that each node i's ranking r_i is drawn independently from some (parameterized) distribution $\rho(r|\boldsymbol{\eta})$ over $\Omega(\mathcal{A})$. We also assume a *ranking distance metric* $d : \Omega(\mathcal{A})^2 \to \mathbb{R}$ which measures similarity between rankings. Finally, a *connection probability function* $c(d) : [0, \infty) \to [0, 1]$ determines the probability that two nodes i, j are connected given the distance $d(r_i, r_j)$ between their rankings. We now detail each of these three components of our model.

Distance Metric on Rankings. We use the "similarity" of two rankings to determine their distance, which will be used below to determine connection probabilities. A variety of well-known distance metrics for rankings can be used [9]. We briefly describe several common distance metrics. A natural set of distances are "d_p distances", where $d_p(r, r') = \sum_{i=1}^{m} |r(a_i) - r'(a_i)|^p$ for $p \in [1, \infty)$. The well-known *footrule* ($p = 1$) and *Spearman* ($p = 2$) distances are instances of this. *Hamming* distance is another natural model, while *Kendall's* τ distance is used, e.g., in psychometrics and social choice, where

$$d_\tau(r, r') = \sum_{k \neq l} I[r(a_k) > r(a_l) \text{ and } r'(a_k) < r'(a_l)]. \tag{7}$$

Here d_τ measures the number of pairwise swaps needed to transform r to r'. As the ranking space $\Omega(\mathcal{A})$ is discrete with finite size $m!$, there are finitely many realizable distances. It is easy to see that, for any $r, r' : 0 \leq d_\tau(r, r') \leq \binom{m}{2}$ and $d_\tau(r, r') \in \mathbb{N} \cup \{0\}$.

Ranking Distributions. The $\rho(r|\boldsymbol{\eta})$ component of our ranking network model accommodates arbitrary ranking distributions. Distributional models of rankings developed in psychometrics and statistics, and now widely used in machine learning and IR [23], include Mallows, Plackett-Luce, Bradley-Terry, and many others (see [28] for an overview). We use the Mallows' ϕ-model in our empirical experiments. It is characterized by a "modal" *reference ranking* σ and a *dispersion parameter* $\phi \in [0, 1)$, with the probability of a ranking r given by $\rho(r|\sigma, \phi) \propto \phi^{d_\tau(r,\sigma)}$.

Connection Probability Function. With ranking-based distance metrics in hand, we adopt standard connection functions for latent-space models (see above). We assume $c(d)$ is integrable and strictly decreasing. We can derive sufficient conditions for connectivity of a ranking network:

Theorem 3.1 *Assume a ranking model $(\rho(r|\boldsymbol{\eta}), c(d|\boldsymbol{\lambda}))$. The induced ranking network is connected with high probability (i.e., with probability $1 - o(1)$ where $o(1) \to 0$ and $n \to \infty$) if*

$$d_M(m) < c^{-1}\left(\frac{\log n}{n}\Big|\boldsymbol{\lambda}\right),$$

where $d_M(m)$ is the maximum possible distance under d given m alternatives.[1]

This can be used to derive suitable conditions for connectivity of specific models. For instance, using the τ distance metric and Waxman connection function $c(x|\alpha, d_0) = \alpha e^{\frac{x}{d_0}}$, the emerging ranking network will be connected with high probability if $\binom{m}{2} < -d_0 \ln \frac{\log n}{\alpha n}$.

The *small world* effect is a commonly observed property of real-world networks: the diameter (or longest shortest path between any pair of nodes) is small, as is the average shortest path length [1].

[1] We assume that m is fixed and independent of the number n of individuals. This is consistent with the applications we have in mind (e.g., social choice, recommender systems, interactive search, etc.). We plan to extend our analysis to the case where m depends on n. Omitted proofs can be found in a longer version of the paper at `http://www.cs.toronto.edu/~abari/papers/RankingNetworks.pdf`

Ranking networks exhibit these properties: the diameter $D(\boldsymbol{\theta})$ and the average shortest path $\langle l(\boldsymbol{\theta})\rangle$ can be approximated and (often) bounded by:

$$D(\boldsymbol{\theta}) \leq \left\lceil \frac{\log(n)}{\log(n-1) + \log c(d_M(m)|\boldsymbol{\lambda})} \right\rceil; \quad \langle l(\boldsymbol{\theta})\rangle \leq \frac{\log(n)}{\log(n-1) + \log c(d_M(m)|\boldsymbol{\lambda})}. \tag{8}$$

These approximations are detailed in the appendix of a longer version of this paper; and our empirical results confirm these properties and the tightness of these bounds. Ranking networks also possess the diameter-shrinking property of real-world networks [21], where diameter (and average shortest path) shrinks as the network grows:

Theorem 3.2 *Fix m and assume $\rho(r|\boldsymbol{\eta})$ distributes probability mass on more than one ranking. The asymptotic diameter of ranking networks is 2 (as $n \to \infty$).*

These and other structural properties of ranking networks can be computed readily if model parameters are given. However, due to the discrete nature of ranking space, computation can be extremely intensive, even for relatively small m, due to the combinatorial size of ranking space. This motivates the development of easy-to-compute approximations for the special class of *distance-based ranking models*, while studying properties that emerge among networks in this class. Our approximation methods also can applied for efficient learning of model parameters.

3.3 Distance-Based Ranking Models

Distance-based ranking distributions [12, 28] have ranking probabilities that decrease with increasing distance from some modal or reference ranking $\sigma \in \Omega(\mathcal{A})$:

$$\rho(r|\sigma,\omega) = \frac{1}{\psi(\omega)} \exp(-\omega d(r,\sigma)), \tag{9}$$

where $\omega \in [0,\infty)$ is a *dispersion parameter* and $\psi(\omega)$ is a normalizing constant. As $\omega \to \infty$, ρ becomes concentrated at the reference ranking σ, whereas for $\omega = 0$, ρ is the uniform distribution. The Mallows ϕ-model above is an example of such a model (with dispersion $\phi = e^{-\omega}$ and distance d_τ). While we focus on unimodal models, mixtures of such models offer additional modeling flexibility [25].

In this section, we assume that ρ (our distance-based ranking distribution) and c (our distance-based connection function) use the same distance metric. For instance, when using d_τ as our distance measure, the ranking distribution is the Mallows ϕ-model and the probability of connection between two nodes is determined based on this same distance.

We first observe that, as $\omega \to \infty$, ranking networks converge on the well-studied ER random graph model $G(n,p)$ with $p = c(0|\boldsymbol{\lambda})$. We can also bound the probability that a node with a ranking r is connected to a randomly chosen node:

Theorem 3.3 *Given reference ranking σ and a distance-based ranking model, for any fixed $\boldsymbol{\theta}$ and any $r \in \Omega(\mathcal{A})$:*

$$\mathcal{D}(\sigma_M,\boldsymbol{\theta}) \leq \mathcal{D}(r,\boldsymbol{\theta}) \leq \mathcal{D}(\sigma,\boldsymbol{\theta}),$$

where σ_M is some ranking at maximum distance from σ.[2]

This theoretical observation has a natural interpretation: individuals that possess more probable or more "popular" rankings have greater odds of connecting with others (hence have higher expected degree). This is not surprising given the nature of distance-based ranking models. In the context of social networks, this is one possible contributing factor to the intuition that people with more "popular" preferences tend to have more friends and social interactions than those with uncommon preferences, i.e., *preference popularity governs social popularity*. From a different perspective, one can use observed node degree to draw inferences about its ranking: higher degree is predictive of more common preferences. (i.e., close to the modal ranking). This can be exploited to support efficient estimation/learning of the reference ranking, the ranking distribution, and preferences of specific individuals. Using Thm. 3.3, it is straightforward to bound $\mathcal{E}(\boldsymbol{\theta})$.

[2] If more than one ranking has maximum distance, one such ranking minimizes $\mathcal{D}$ (not necessarily all).

Proposition 3.4 *Given a distance-based ranking model, $\mathcal{E}(\boldsymbol{\theta})$ is bounded by*

$$\mathcal{D}(\sigma_M, \boldsymbol{\theta}) \leq \mathcal{E}(\boldsymbol{\theta}) \leq \mathcal{D}(\sigma, \boldsymbol{\theta}),$$

where σ_M is some ranking at maximum distance from σ.

The upper bound $\mathcal{D}(\sigma, \boldsymbol{\theta})$ can be computed efficiently (in $O(m^2)$ time):

$$\mathcal{D}(\sigma, \boldsymbol{\theta}) = \hat{\mathcal{D}}(m, \omega, \boldsymbol{\eta}) = \frac{1}{\psi(\omega)} \sum_{k=0}^{d_M(m)} n_k e^{-\omega k} c(k|\boldsymbol{\lambda}), \tag{10}$$

where $d_M(m)$ is the maximum possible distance for given d when there are m alternatives (e.g., for Kendall-τ, $d_M(m) = \binom{m}{2}$) and n_i is the number of rankings of distance i from an arbitrary fixed ranking. We can efficiently compute n_i for any d either in closed form or using dynamic programming. For instance, n_i can be computed (once) for Kendall's τ in $O(m^3)$ time via:

$$T(j,k) = \begin{cases} 1, & k = 0 \text{ and } j \geq 1 \\ 0, & k > \binom{m}{2} \\ T(j, k-1) + T(j-1, k), & k \leq j-1 \\ T(j, k-1) + T(j-1, k) - T(j-1, k-j), & k > j-1 \end{cases} \tag{11}$$

If d is symmetric in the sense that $n_i = n_{d_M(m)-i}$ for all $i \leq d_M(m)$, a lower bound $\mathcal{D}(\sigma_M, \boldsymbol{\theta})$ can be computed in $O(m^2)$ time:

$$\mathcal{D}(\sigma_M, \boldsymbol{\theta}) = \check{\mathcal{D}}(m, \omega, \boldsymbol{\lambda}) = \frac{1}{\psi(\omega)} \sum_{k=0}^{d_M(m)} n_k e^{-\omega k} c(d_M(m) - k|\boldsymbol{\lambda}) \tag{12}$$

We have empirically observed that $\mathcal{D}(r, \boldsymbol{\theta})$ (usually) decreases as $d(r, \sigma)$ increases. In other words, the distance of an individual's ranking to the reference ranking is negatively correlated with its degree. Define the linear function

$$\tilde{\mathcal{D}}(d, \boldsymbol{\theta}) = \left(1 - \frac{d}{d_M(m)}\right) \mathcal{D}(\sigma, \boldsymbol{\theta}) + \left(\frac{d}{d_M(m)}\right) \mathcal{D}(\sigma_M, \boldsymbol{\theta}). \tag{13}$$

We can approximate $\mathcal{D}(r, \boldsymbol{\theta})$ by $\tilde{\mathcal{D}}(d(r, \sigma), \boldsymbol{\theta})$, which can be used to effectively approximately other structural network properties. For example, the edge density $\mathcal{E}$ and degree distribution P, respectively, be approximated by:

$$\tilde{\mathcal{E}}(\boldsymbol{\theta}) = \frac{1}{\psi(\omega)} \sum_{k=0}^{d_M(m)} n_k e^{-\omega k} \tilde{\mathcal{D}}(k, \boldsymbol{\theta}), \text{ and} \tag{14}$$

$$\tilde{P}(k, \boldsymbol{\theta}) = \frac{\binom{n-1}{k}}{\psi(\omega)} \sum_{i=0}^{d_M(m)} n_i e^{-\omega i} \tilde{\mathcal{D}}(i, \boldsymbol{\theta})^k \left(1 - \tilde{\mathcal{D}}(i, \boldsymbol{\theta})\right)^{n-1-k}. \tag{15}$$

By pre-computing n_i values, these functions can be computed in $O(m^2)$ time (cf. the $O((m!)^2)$ time required for naive exact computation). For large m, even this might problematic; but if $c(\cdot, \boldsymbol{\lambda})$ is convex, one can (loosely) approximate $\mathcal{D}$ and edge density $\mathcal{E}$ in $O(1)$ time by:

$$\tilde{\hat{\mathcal{D}}}(x, \boldsymbol{\theta}) = \left(1 - \frac{x}{d_M(m)}\right) c\left(\frac{d_M(m) e^{-\omega}}{1 + e^{-\omega}}\right) + \frac{x}{d_M(m)} c\left(\frac{d_M(m)}{1 + e^{-\omega}}\right), \text{ and} \tag{16}$$

$$\tilde{\hat{\mathcal{E}}}(\boldsymbol{\theta}) = \frac{1}{1 + e^{-\omega}} c\left(\frac{d_M(m) e^{-\omega}}{1 + e^{-\omega}}\right) + \frac{e^{-\omega}}{1 + e^{-\omega}} c\left(\frac{d_M(m)}{1 + e^{-\omega}}\right). \tag{17}$$

Our approximations $\tilde{\mathcal{E}}(\boldsymbol{\theta})$ and $\tilde{\mathcal{D}}(d(r, \sigma), \boldsymbol{\theta})$ can also be exploited for efficient model learning. Exact evaluation of the likelihood function of a ranking network when dealing with missing data (i.e., when some node rankings are unobserved) has computation of edge density $\mathcal{E}(\boldsymbol{\theta})$ and average connection probability $\mathcal{D}(r, \boldsymbol{\theta})$ as its main bottlenecks. Of course, one can learn model parameters using methods iterative methods such as EM, but direct evaluation of (an approximation of) the likelihood function using these approximations may be support more efficient learning and inference (preliminary tests seem to support this, but this is the topic of ongoing research).

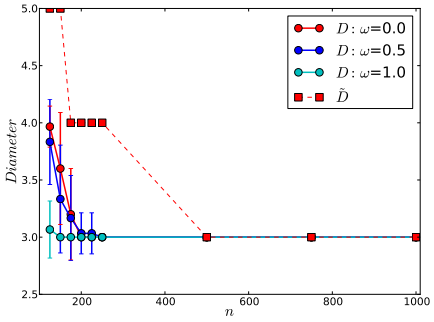

Figure 1: Mean and std. dev. of observed diameters (30 instances) for various ω and n, and its approximation $\tilde{D}$ ($m = 3$, $\alpha = 2$, $\gamma = 0.8$, and $\beta = 2$).

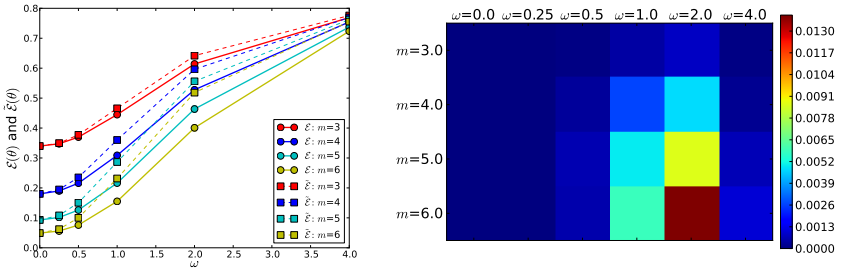

Figure 2: (a) Mean observed edge density $\mathcal{E}(\boldsymbol{\theta})$ (30 instances) and its approximation $\tilde{\mathcal{E}}(\boldsymbol{\theta})$; (b) mean squared error of approximate edge density; varying number of options m, dispersion ω ($n = 1000$, $\alpha = 2$, $\beta = 2$, $\gamma = 0.8$).

4 Numerical Case Study with Mallows Models

We describe experiments on a version of our ranking network model, using the Mallows ϕ-model under various parameter settings. We compare various structural properties (e.g., degree distribution, edge density, diameter) of the resulting networks with the predictions of our approximation methods.

Experimental Setup. We use a variant of the connection probability of [33]:

$$c(d|\boldsymbol{\lambda}) = \gamma \left(1 + \frac{d}{\beta}\right)^{-\alpha}. \tag{18}$$

Here β controls average degree and $\alpha > 1$ determines the extent of homophily. We introduce $\gamma \in (0, 1]$ to control the probability of connecting nodes with the same ranking (to account for the discrete nature of ranking space). Unless noted, all experiments are run on networks with $n = 1000$ nodes, $\alpha = 2$, $\beta = 2$ and $\gamma = 0.8$, while varying m. For each parameter setting, we report results over 30 random ranking networks.

Diameter. We first examine the effect of n and ω on the diameter of the emergent ranking networks, and compare observed diameter to our approximation $\tilde{D}$ (see Eq. 8). We fix $m = 5$. Fig. 1 confirms that diameter shrinks as n increases. Unsurprisingly, diameter decreases with increasing ω (i.e., more uniform distributions give larger diameter), largely due to the increased edge density caused by increasing ω (see below). $\tilde{D}$ provides a reasonable upper bound of diameter for any value of ω, with bounds that are very tight when n reaches 500.

Edge Density. The effect of m and ω on edge density is illustrated in Fig. 2. Fig. 2(a) demonstrates that edge density increases with ω but decreases with m (compare boxes on the solid lines). Error bars are very tight (and barely observable). Our approximation $\tilde{\mathcal{E}}$ (Eq. 14) is relatively close to the observed edge density, especially when ω is relatively small ($\omega \leq 0.5$) or large ($\omega \geq 4$).

To better quantify the accuracy of our approximation, we compute the mean squared error (MSE) between the actual edge density and our approximation over 30 randomly generated ranking networks. Fig. 2(b) shows that MSE is relatively low with a maximum of roughly 0.014 for $m = 6$, $\omega = 2$. MSE first increases and then decreases with ω. Moreover, MSE increases with m especially

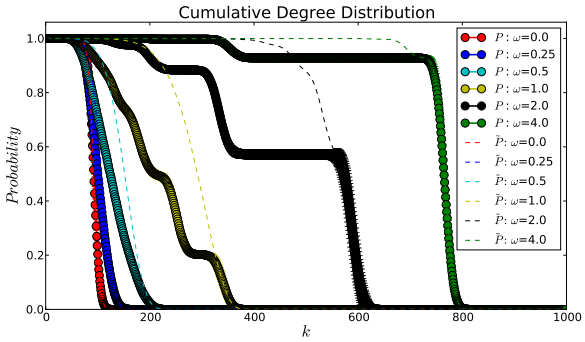

Figure 3: The mean cumulative degree distribution (30 instances) and its approximation, varying ω ($m = 5$, $n = 1000$, $\alpha = 2$, $\beta = 2$, $\gamma = 0.8$).

for $\omega = 1, 2$. This suggests that, for large m and small ω, our edge density approximation should be used cautiously. Refined approximations that account for the value of ω are being explored.

Degree Distribution. Fig. 3 shows the impact of ω on the degree distribution of ranking networks under the Mallows model, as well as the effectiveness of our approximation $\tilde{P}$ (Eq. 15). The figure shows *cumulative degree distribution* to reduce noise in the plots: given degree distribution $P(k, n)$, the cumulative degree distribution is $P_c(k, n) = \sum_{l=k}^{n} P(l, n)$ (i.e., probability that a node n has the degree $d \geq k$). The mean cumulative degree distribution is shown for various values of ω with $m = 5$ and $n = 1000$. The expected degree increases with ω; e.g., with $\omega = 0$, most of nodes have degree around 100, whereas for $\omega = 4$, most nodes have degree around 800 (since mean degrees is $n\mathcal{E}$). However, variance in the degree distribution initially increases as ω increases from 0 but then decreases. For example, variance is low at $\omega = 0$ and $\omega = 4$, but high for $\omega = 1$ and $\omega = 2$. Interestingly, with $\omega = 1$ and $\omega = 2$, the the cumulative degree distribution shows several distinct modes in the degree distribution. The relation between these modes and community structure in ranking networks is something that remains to be explored.

Fig. 3 also shows that the approximate degree distribution $\tilde{P}$ is reasonably close to actual observed degree distribution when $\omega \leq 0.5$ and $\omega \geq 4$, though it fails to account for the distinct modes, especially for $\omega = 1, 2$. Once again, more accurate approximations (sensitive to ω) may be possible.

5 Concluding Remarks and Future Work

We have introduced *ranking networks*, a class of attribute-based (or latent-space) network formation models in which node attributes are rankings over a set of options, and connections are formed between nodes based on the similarity of their underlying rankings. We studied some structural properties (e.g., diameter, connectivity, edge density) of these networks, and showed that our model possesses some characteristics of real-world networks (e.g., shrinking diameter).

This model is a starting point for the broader investigation of the impact of rankings, and preferences in particular, on network formation. Future directions include: the analysis of more realistic ranking distributions (e.g., mixture models); extensions of our model that account for heterogeneity/heterophily; and modeling the dynamics and mutability of the underlying rankings themselves in response to network connections. Of practical importance is studying the extent to which rankings, such as preferences, play a role in shaping connections in real-world networks, and how the induced correlations can be best exploited in applications such as recommender systems, advertising, social choice and voting, web search and information retrieval.

References

[1] Réka Albert and Albert-László Barabási. Statistical mechanics of complex networks. *Reviews of modern physics*, 74(1):47, 2002.

[2] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

[3] Alain Barrat, Marc Barthelemy, and Alessandro Vespignani. *Dynamical Processes on Complex Networks*. Cambridge University Press, 2008.

[4] Marc Barthélemy. Spatial networks. *Physics Reports*, 499(1):1–101, 2011.

[5] Mustafa Bilgic, Lilyana Mihalkova, and Lise Getoor. Active learning for networked data. In *Proceedings of the 27th International Conference on Machine Learning*, pages 79–86, 2010.

[6] Marián Boguñá and Romualdo Pastor-Satorras. Class of correlated random networks with hidden variables. *Physical Review E*, 68:036112, Sep 2003.

[7] Marián Boguñá, Romualdo Pastor-Satorras, Albert Díaz-Guilera, and Alex Arenas. Models of social networks based on social distance attachment. *Physical Review E*, 70:056122, Nov 2004.

[8] Béla Bollobás. *Random graphs*. Cambridge University Press, 2001.

[9] Persi Diaconis. *Group Representations in Probability and Statistics*. Institute of Mathematical Statistics, Hayward, California, 1988.

[10] David A. Easley and Jon M. Kleinberg. *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. Cambridge University Press, Cambridge, 2010.

[11] Paul Erdős and Alfréd Rényi. On random graphs. *Publicationes Mathematicae Debrecen*, 6:290–297, 1959.

[12] Michael A Fligner and Joseph S Verducci. Distance based ranking models. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 359–369, 1986.

[13] Lise Getoor and Christopher P. Diehl. Link mining: A survey. *ACM SIGKDD Explorations Newsletter*, 7(2):3–12, December 2005.

[14] Peter D. Hoff, Adrian E. Raftery, and Mark S. Handcock. Latent space approaches to social network analysis. *Journal of the American Statistical Association*, 97(460):1090–1098, 2002.

[15] Matthew O. Jackson. *Social and Economic Networks*. Princeton University Press, Princeton, 2008.

[16] Amin Karbasi, Stratis Ioannidis, and Laurent Massoulié. Content search through comparisons. In *Automata, Languages and Programming*, pages 601–612. Springer, 2011.

[17] Myunghwan Kim and Jure Leskovec. Modeling social networks with node attributes using the multiplicative attribute graph model. In *UAI'2011*, pages 400–409.

[18] Myunghwan Kim and Jure Leskovec. Latent multi-group membership graph model. In *Proceedings of the 29th International Conference on Machine Learning*, 2012.

[19] Myunghwan Kim and Jure Leskovec. Multiplicative attribute graph model of real-world networks. *Internet Mathematics*, 8(1-2):113–160, 2012.

[20] Jure Leskovec, Deepayan Chakrabarti, Jon Kleinberg, Christos Faloutsos, and Zoubin Ghahramani. Kronecker graphs: An approach to modeling networks. *JMLR*, 11:985–1042, 2010.

[21] Jure Leskovec, Jon Kleinberg, and Christos Faloutsos. Graphs over time: Densification laws, shrinking diameters and possible explanations. In *KDD'05*, pages 177–187, 2005.

[22] David Liben-Nowell and Jon Kleinberg. The link-prediction problem for social networks. *Journal of the American Society for Information Science and Technology*, 58(7):1019–1031, 2007.

[23] Tie-Yan Liu. Learning to rank for information retrieval. *Foundations and Trends in Information Retrieval*, 3(3):225–331, 2009.

[24] Linyuan Lü and Tao Zhou. Link prediction in complex networks: A survey. *Physica A: Statistical Mechanics and its Applications*, 390(6):1150–1170, 2011.

[25] Tyler Lu and Craig Boutilier. Learning mallows models with pairwise preferences. In *Proceedings of the 28th International Conference on Machine Learning*, pages 145–152, 2011.

[26] Sofus A. Macskassy. Using graph-based metrics with empirical risk minimization to speed up active learning on networked data. In *KDD'09*, pages 597–606.

[27] Colin L. Mallows. Non-null ranking models. In *Biometrika*, volume 44, pages 114–130, 1957.

[28] John I. Marden. *Analyzing and Modeling Rank Data*. Chapman and Hall, London, 1995.

[29] Mark Newman. *Networks: An Introduction*. Oxford University Press, 2010.

[30] Mark Newman, Albert-László Barabási, and Duncan J. Watts. *The Structure and Dynamics of Networks*. Princeton University Press, 2006.

[31] Mathew Penrose. *Random Geometric Graphs*. Oxford University Press, USA, 2003.

[32] Prithviraj Sen, Galileo Mark Namata, Mustafa Bilgic, Lise Getoor, Brian Gallagher, and Tina Eliassi-Rad. Collective classification in network data. *AI Magazine*, 29(3):93–106, 2008.

[33] M. Ángeles Serrano, Dmitri Krioukov, and Marián Boguñá. Self-similarity of complex networks and hidden metric spaces. *Physical Review Letters*, 100:078701, Feb 2008.

[34] Duncan J. Watts and Steven H. Strogatz. Collective dynamics of small-world networks. *Nature*, 393(6684):440–442, 1998.

[35] Bernard M. Waxman. Routing of multipoint connections. *Selected Areas in Communications, IEEE Journal on*, 6(9):1617–1622, December 1988.

[36] Stephen J. Young and Edward R. Scheinerman. Random dot product graph models for social networks. In *WAW'07*, pages 138–149, 2007.

This Appendix first presents proofs of theoretical analyses in Section A. Then, Section B discusses the rational behind the proposed approximations. Finally, we discuss the details of dynamic programming methods used to compute the number of rankings with a specific Kendall-distance from a fixed ranking in Section C.

A Proofs

Proof of Theorem 3.1 We note that the connection probability between two nodes with rankings r and $r^{'}$ can be written as $c(d(r,r^{'}),\boldsymbol{\lambda}) = c(d_M(m),\boldsymbol{\lambda}) + (c(d(r,r^{'}),\boldsymbol{\lambda}) - c(d_M(m),\boldsymbol{\lambda}))$. So we can consider that the network formation – under ranking network model – occurs in two phases. First any two nodes get connected to each other with fixed probability of $c(d_M(m),\boldsymbol{\lambda})$ (similar to random graph model $G(n,p)$ with $p = c(d_M(m),\boldsymbol{\lambda})$). Then, any pair of nodes with ranking r and $r^{'}$ get connected with the probability $c(d(r,r^{'}),\boldsymbol{\lambda}) - c(d_M(m),\boldsymbol{\lambda})$. So if we find some conditions under which the network generated after phase 1 is almost surely connected, over the second phase, the connectivity property should holds since only some other edges might be added to the network without removal of any existing edge.

It is well-studied that the random graph $G(n,p)$ is almost surely connected when $p > \frac{\log n}{n}$ [8]. Since the first phase is actually the random graph model $G(n,p)$ with $p = c(d_M(m),\boldsymbol{\lambda})$. So we require

$$c(d_M(m)|\boldsymbol{\lambda}) > \frac{\log n}{n}$$

Since $c(.|\boldsymbol{\lambda})$ is a strictly decreasing function, we have

$$d_M(m) < c^{-1}\left(\frac{\log n}{n}\Big|\boldsymbol{\lambda}\right) \quad \blacksquare$$

Proof of Theorem 3.2 Our proof has following steps: first we will show that asymptotic diameter should be at most 2 by demonstrating that for any two arbitrary nodes are connected through at least another nodes in the network. Second, we show that each node can not be connected to all other nodes in the network, thus resulting the asymptotic diameter of more than 1. Then, the asymptotic diameter has to be 2.

For the first step, we start by the observation that connection probability of any two nodes regardless of their rankings is at least $p = c(d_M(m)|\boldsymbol{\lambda})$. We note that $c(d_M(m)|\boldsymbol{\lambda}) > 0$ for any distance metric d and $m \geq 2$; so $p > 0$. Let denote the probability that i and j has not common neighbour k by $P(not\ i \sim k \sim j)$. Similarly let $P(not\ i \sim S \sim j)$ denote the probability that i and j has not any element of set S as a common neighbour. It is straight forward to see that $P(not\ i \sim S \sim j) \leq 1-p^2$. Similarly, due to independence of connections, one can observe that $P(not\ i \sim \mathcal{N} \setminus \{i,j\} \sim j) \leq (1-p^2)^{n-2}$. Since $p > 0$, $\lim_{n\to\infty} P(not\ i \sim \mathcal{N} \setminus \{i,j\} \sim j) = 0$. This implies that i and j have to have at least one common neighbour when $n \to \infty$. So the shortest path between these two nodes have at least length of 2. As i and j where two arbitrary nodes, any two pair of nodes at least have the shortest path of length 2. So the asymptotic diameter of the network is at most 2.

We now show that asymptotic diameter is greater than 1. We prove this by contradiction. We assume that the asymptotic diameter is 1. As the mass probability is not only on one ranking, so there are at least two rankings r_m and r_M such that $0 < \rho(r_m) \leq \rho(r_M)$ and also $\rho(r_m) \leq \rho(r) \leq \rho(r_M)$ for any r with $\rho(r) > 0$. Let n_{r_m} and n_{r_M} denote the number of nodes with rankings r_m and r_M when there are n nodes. We consider sufficiently large n such that $n_{r_m}, n_{r_M} > 0$. As r_m and r_M are distinct, $d(r_m,r_M) > 0$. So, $c(d(r_m,r_M)) < c(0)$. Let $P(r_m,n_{r_M})$ denote the probability that one node with r_m ranking get connected to all nodes with r_M rankings. It is straightforward to see that $P(r_m,n_{r_M}) = c(d(r_m,r_M))^{n_{r_M}}$. We note when $n \to \infty$, $n_{r_M} \to \infty$ when m is fixed. As $c(d(r_m,r_M)) < c(0) \leq 1$, $\lim_{n\to\infty} P(r_m,n_{r_M}) = 0$. So there should be some nodes with r_M that a node with r_m is not directly connected to. Therefore, the asymptotic diameter can not be 1. Based on this and the first step proof that asymptotic diameter is ≤ 2. The asymptotic diameter must be 2. ∎

Proof of Theorem 3.3. We first note that for $\omega = 0$, the proof is trivial: When $\omega = 0$, for any $r \in \Omega(\mathcal{A})$, $\rho(r|\boldsymbol{\theta}) = \frac{1}{m!}$. Hence,

$$D(r,\boldsymbol{\theta}) = \sum_{r^{'}\in\Omega(\mathcal{A})} \rho(r^{'}|\boldsymbol{\theta}) c\left(d_\tau(r^{'},r)|\boldsymbol{\theta}\right) = \frac{1}{m!}\sum_{r^{'}\in\Omega(\mathcal{A})} c\left(d_\tau(r^{'},r)|\boldsymbol{\theta}\right)$$
$$= \frac{1}{m!}\sum_{i=0}^{d_M(m)} n_i c\left(i|\boldsymbol{\theta}\right),$$

where n_i represent the number of rankings with the distance i from a specific ranking and $d_M(m)$ is the maximum possible distance under deployed distance metric d. We note that the value of $D(r,\boldsymbol{\theta})$ is same for any $r \in \Omega(\mathcal{A})$. So, for this case, for any $r \in \Omega(\mathcal{A})$, we have

$$\mathcal{D}(\sigma_M,\boldsymbol{\theta}) = \mathcal{D}(r,\boldsymbol{\theta}) = \mathcal{D}(\sigma,\boldsymbol{\theta}),$$

where σ_M is a ranking with maximum possible distance to σ.

To prove the inequities for the case $\omega \in (0,\infty)$, we first let $k_{ij}^{(r)}$ denote the number of rankings with the distance of i to the reference ranking σ and the distance of j to an arbitrary ranking r. Once again, let n_i represent the number of rankings with the the distance i from a specific fixed ranking, where $i \in \{0,\cdots,d_M(m)\}$. So, we can write

$$\mathcal{D}(r,\boldsymbol{\theta}) = \sum_{r^{'}\in\Omega(\mathcal{A})} \rho(r^{'}|\boldsymbol{\theta}) c\left(d(r^{'},r)|\boldsymbol{\theta}\right)$$
$$= \frac{1}{\psi(\omega)}\sum_{i=0}^{d_M}\sum_{j=0}^{d_M} k_{ij}^{(r)} e^{-\omega i} c(j|\boldsymbol{\theta}). \tag{19}$$

We also observe that for any fixed $r \in \Omega(\mathcal{A})$,

$$\forall j \in \{0,\ldots,d_M\}, \sum_{i=0}^{d_M} k_{ij}^{(r)} = n_j \quad \text{and} \quad \forall i \in \{0,\ldots,d_M\}, \sum_{j=0}^{d_M} k_{ij}^{(r)} = n_i. \tag{20}$$

We first focus on proving $\mathcal{D}(r,\boldsymbol{\theta}) \leq \mathcal{D}(\sigma,\boldsymbol{\theta})$ side of the inequality. By using Eq. 19, we can view $\mathcal{D}(r,\boldsymbol{\theta})$ as a function over matrix $\mathbf{K}^{(r)}$, denoted by $f\left(\mathbf{K}^{(r)}\right)$, with the constraints stated in Eq. 20. Now, we explore for which feasible instance of $\mathbf{K}^{(r)}$, $f(\mathbf{K}^{(r)})$ is maximized. Using Lemma A.1, it follows that $f\left(\mathbf{K}^{(r)}\right)$ is maximized with a diagonal matrix $\mathbf{K}^{(r^*)}$ with $k_{ii}^{(r^*)} = n_i$ for all $i \in \{0,\ldots,d_M\}$. This implies that $k_{00}^{r^*} = n_0 = 1$; so there is one ranking, say $r^{'}$, which has the distance of 0 to both r^* and σ. Obviously, this can be only true when $r^* = r^{'} = \sigma$ (based on identity of indiscernibles property of metric distance). It follows that σ maximize the function $\mathcal{D}(r,\boldsymbol{\theta})$ when $\boldsymbol{\theta}$ is fixed. Hence, for any $r \in \Omega(\mathcal{A})$: $\mathcal{D}(r,\boldsymbol{\theta}) \leq \mathcal{D}(\sigma,\boldsymbol{\theta})$.

We now focus on $\mathcal{D}(\sigma_M,\boldsymbol{\theta}) \leq \mathcal{D}(r,\boldsymbol{\theta})$ side of the inequality. By using Eq. 19, we can view $\mathcal{D}(r,\boldsymbol{\theta})$ as a function over matrix $\mathbf{K}^{(r)}$, denoted by $f\left(\mathbf{K}^{(r)}\right)$, with the constraints stated in Eq. 20. Using Lemma A.3, it follows that r^* which minimizes $f\left(\mathbf{K}^{(r)}\right)$ must have distance $d_M(m)$ to the reference ranking. So $r^* = \sigma_M$ must be one of those rankings with the maximum distance to σ. ∎

Lemma A.1 *Assume* $f(\mathbf{X}) = \sum_{i=0}^{d}\sum_{j=0}^{d} a_{ij}x_{ij}$ *and* $a_{ij} = \frac{1}{\psi(\omega)}e^{-\omega i}c(j)$ *where* $\omega \in (0,\infty)$ *and* $c(x|\boldsymbol{\lambda})$ *is a decreasing connection probability function. Consider this constraint maximization problem:*

$$\begin{aligned}
\max_{\mathbf{X}\in\mathbb{R}^{(d+1)\times(d+1)}} \quad & f(\mathbf{X}) \\
\textit{subject to:} \quad & \forall i \in \{0,\ldots,d\},\ \sum_{j=0}^{d} x_{ij} = c_i \\
& \forall j \in \{0,\ldots,d\},\ \sum_{i=0}^{d} x_{ij} = c_j \\
& \forall i,j \in \{0,\ldots,d\},\ x_{ij} \geq 0
\end{aligned} \tag{21}$$

where $\forall i$, c_i *is a positive constant and given. The solution to this maximization problem is the diagonal matrix* $\mathbf{X}^*$ *such that* $\forall i$, $x_{ii}^* = c_i$.

Proof of Lemma A.1. We prove by contradiction. The proof strategy is as follows. We assume that $\mathbf{X}^*$ maximizes the objective function $f(.)$ but is not diagonal. Then, by slight modification of some elements in $\mathbf{X}^*$, we create matrix $\mathbf{Y}^*$ and show that it is indeed feasible solution. Afterwards, we show that $f(\mathbf{Y}^*) > f(\mathbf{X}^*)$ which contacts the maximality of $\mathbf{X}^*$.

Let $\mathbf{X}^*$ maximizes the objective function $f(.)$ and assume it is not diagonal. Consider this decomposition $\mathbf{X}^* = \mathbf{L}^* + \mathbf{D}^* + \mathbf{U}^*$ where $\mathbf{L}^*$, $\mathbf{D}^*$, and $\mathbf{U}^*$ are strictly lower triangular, diagonal, and strictly upper triangular matrices respectively. Since $\mathbf{X}^*$ is not diagonal, at least, one of $\mathbf{L}^*$ and $\mathbf{U}^*$ must have none-zero elements. Moreover, from Lemma A.2, we can conclude that both $\mathbf{L}^* \neq \mathbf{0}$ and $\mathbf{U}^* \neq \mathbf{0}$. We can define none-empty set $S = \{(i,j) | u^*_{ij} > 0 \; and \; j > i\}$. We consider the lexicographical ordering on S such that

$$(i,j) < (k,l) \Leftrightarrow i < k \; or \; (i = k \; and \; j < k).$$

There should be (i_m, j_m) which is the minimum element in S based on our lexicographical ordering. Based on the definition of S and the minimality of (i_m, j_m) in S, we observe that $u^*_{i_m j_m} > 0$ and $u^*_{k i_m} = 0$ for any $k < i_m$. Hence,

$$\sum_{k<i_m} x^*_{k i_m} = 0 \;\; and \;\; x^*_{i_m j_m} > 0. \tag{22}$$

From two constraints (see Eq. 21) related two row i_m and column i_m, we can write:

$$\sum_j x^*_{i_m j} = \sum_k x^*_{k i_m} = c_{i_m} \implies \sum_j x^*_{i_m j} - \sum_k x^*_{k i_m} = 0 \implies$$

$$\sum_{j \neq i_m} x^*_{i_m j} - \sum_{k \neq i_m} x^*_{k i_m} = 0 \implies \sum_{j \neq i_m} x^*_{i_m j} - \left(\sum_{k<i_m} x^*_{k i_m} + \sum_{k>i_m} x^*_{k i_m} \right) = 0$$

$$\implies x^*_{i_m j_m} - \left(\sum_{k<j_m} x^*_{k i_m} + \sum_{k>j_m} x^*_{k i_m} \right) \leq 0$$

Using this and Eq. 25, it follows that

$$0 < x^*_{i_m j_m} \leq \sum_{k>i_m} x^*_{k i_m} \implies \sum_{k>i_m} x^*_{k i_m} > 0.$$

Hence, there is (at least) $k > i_m$ such that $x^*_{k i_m} > 0$. We can define

$$\delta = \min(x^*_{k i_m}, x^*_{i_m, j_m}).$$

Note that $\delta > 0$ since both $x^*_{k i_m}$ and $x^*_{i_m, j_m}$ are positive. Let

$$B = \{(i_m, i_m), (i_m, j_m), (k, i_m), (k, j_m)\},$$

using δ, we can define matrix $\mathbf{Y}^*$ out of $\mathbf{X}^*$ as follows:

$$y^*_{ij} = \begin{cases} x^*_{ij}, & (i,j) \notin B \\ x^*_{i_m i_m} + \delta, & (i,j) = (i_m, i_m) \\ x^*_{i_m j_m} - \delta, & (i,j) = (i_m, j_m) \\ x^*_{k i_m} - \delta, & (i,j) = (k, i_m) \\ x^*_{k j_m} + \delta, & (i,j) = (k, j_m) \end{cases} \tag{23}$$

It is almost straightforward to see that $\mathbf{Y}^*$ satisfies all constraints stated in Eq. 21. We can first observe that $y^*_{ij} \geq 0$ for all $(i,j) \in B$ mostly because of the way that δ is defined and $\delta > 0$. Moreover, $y^*_{ij} = x^*_{ij} \geq 0$ for all $(i,j) \notin B$. Since $\mathbf{Y}^*$ is different than $\mathbf{X}^*$ only in four corners of a box determined by coordinates in B, to show that $\mathbf{Y}^*$ satisfy all other constraints, it is only sufficient to demonstrate the constraints related to rows i_m and k, and columns i_m and j_m are satisfied (note that as the other rows and columns are unchanged compared to $\mathbf{X}^*$, their corresponding constraints should be already satisfied.). From definition of $\mathbf{Y}^*$ in Eq. 23, we observe that whenever δ is added to the original X_{ij}, it is subtracted from another element in the same row and another element in the same column; this has guaranteed that the constraints are still met.

As $\mathbf{Y}^*$ is a feasible solution, we can write:

$$\begin{aligned} f(\mathbf{Y}^*) - f(\mathbf{X}^*) &= \sum_i \sum_j y^*_{ij} a_{ij} - \sum_i \sum_j x^*_{ij} a_{ij} = \sum_{(i,j)\in B} a_{ij}(y^*_{ij} - x^*_{ij}) \\ &= a_{i_m i_m}\delta - a_{i_m j_m}\delta - a_{k i_m}\delta + a_{k j_m}\delta \\ &= \frac{\delta}{\psi(\omega)}\left[\left(e^{-\omega i_m} c(i_m) - e^{-\omega i_m} c(j_m)\right) - \left(e^{-\omega k} c(i_m) - e^{-\omega k} c(j_m)\right)\right] \\ &= \frac{\delta}{\psi(\omega)}\left(e^{-\omega i_m} - e^{-\omega k}\right)\left(c(i_m) - c(j_m)\right) \end{aligned}$$

Since $j_m > i_m$ and $c(.)$ is decreasing function, $c(i_m) - c(j_m) > 0$. Also, since $i_m < k$ and $\omega \in (0, \infty)$, we have $e^{-\omega i_m} - e^{-\omega k} > 0$. Hence,

$$\frac{\delta}{\psi(\omega)}\left(e^{-\omega i_m} - e^{-\omega k}\right)\left(c(i_m) - c(j_m)\right) > 0 \implies f(\mathbf{Y}^*) > f(\mathbf{X}^*),$$

which contradicts the optimality of $\mathbf{X}^*$. So $\mathbf{X}^*$ has to be diagonal. Using diagonal property of $\mathbf{X}^*$ and constraints stated in Eq. 21, we have $x^*_{ii} = c_i$ and for each $i \neq j \in \{1, \cdots, d\}$, $x^*_{ij} = 0$. ∎

Lemma A.2 *Assume matrix* $\mathbf{X}$ *satisfies the constraints in Eq. 21 and decomposes into diagonal matrix* $\mathbf{D}$*, strictly upper triangular matrix* $\mathbf{U}$ *and strictly lower triangular matrix* $\mathbf{L}$ *(i.e.,* $\mathbf{X} = \mathbf{L} + \mathbf{D} + \mathbf{U}$*):*

$$\mathbf{L} \neq \mathbf{0} \iff \mathbf{U} \neq \mathbf{0}$$

Proof of Lemma A.2 We first prove the $\mathbf{L} \neq \mathbf{0} \implies \mathbf{U} \neq \mathbf{0}$ direction and then the other direction. If $\mathbf{L} \neq \mathbf{0}$ then there exist (at least) one (i, j) such that $l_{ij} > 0$ and $i > j$ (note that $l_{ij} = 0$ for all $j \geq i$ based on the strictly lower triangular property of $\mathbf{L}$). Hence, we can define the non-empty set $S = \{(j, i) | l_{ij} > 0 \; and \; i > j\}$. As S has finite number of elements, we can order elements of S lexicographically such that $(j, i) > (k, l) \; if \; and \; only \; if \; j > k \; or \; (j = k \; and \; i > l)$. Let (j_m, i_m) be the minimum element in S based on our lexicographical ordering. Based on the definition of S and the minimality of (j_m, i_m) in S, we observe that $l_{i_m j_m} > 0$ and $l_{j_m k} = 0$ for any $k < j_m$. Hence,

$$\sum_{k<j_m} x_{j_m k} = 0 \;\; and \;\; x_{i_m j_m} > 0. \tag{24}$$

From constraints stated in Eq. 21, we have:

$$\begin{aligned} &\sum_i x_{i j_m} = \sum_k x_{j_m k} = c_{j_m} \implies \sum_i x_{i j_m} - \sum_k x_{j_m k} = 0 \implies \\ &\sum_{i \neq j_m} x_{i j_m} - \sum_{k \neq j_m} x_{j_m k} = 0 \implies \sum_{i \neq j_m} x_{i j_m} - \left(\sum_{k<j_m} x_{j_m k} + \sum_{k>j_m} x_{j_m k}\right) = 0 \\ &\implies x_{i_m j_m} - \left(\sum_{k<j_m} x_{j_m k} + \sum_{k>j_m} x_{j_m k}\right) \leq 0 \end{aligned}$$

Using this and Eq. 24, it follows that

$$0 < x_{i_m j_m} \leq \sum_{k>j_m} x_{j_m k} \implies \sum_{k>j_m} x_{j_m k} > 0.$$

Hence, there is (at least) $l > j_m$ such that $x_{j_m l} > 0$. So, $u_{j_m l} > 0 \implies \mathbf{U} \neq \mathbf{0}$.

The similar type of argument can be made for the other direction of $\mathbf{U} \neq \mathbf{0} \implies \mathbf{L} \neq \mathbf{0}$. If $\mathbf{U} \neq \mathbf{0}$ then there exist (at least) one (i, j) such that $u_{ij} > 0$ and $i < j$ (note that $u_{ij} = 0$ for all $j \leq i$ based on the strictly upper triangular property of $\mathbf{U}$). Hence, we can define the non-empty set $T = \{(i, j) | u_{ij} > 0 \; and \; i < j\}$. As T has finite number of elements, we can order elements of T lexicographically such that $(i, j) > (k, l) \; if \; and \; only \; if \; i > k \; or \; (i = k \; and \; j > l)$. Let (i_m, j_m) be the minimum element in T based on our lexicographical ordering. Based on the definition of T

and the minimality of (i_m, j_m) in T, we observe that $u_{i_m j_m} > 0$ and $u_{k i_m} = 0$ for any $k < i_m$. Hence,

$$\sum_{k<i_m} x_{k i_m} = 0 \ \ and \ \ x_{i_m j_m} > 0. \tag{25}$$

From constraints stated in Eq. 21, we have:

$$\sum_j x_{i_m j} = \sum_k x_{k i_m} = c_{i_m} \implies \sum_j x_{i_m j} - \sum_k x_{k i_m} = 0 \implies$$

$$\sum_{j \neq i_m} x_{i_m j} - \sum_{k \neq i_m} x_{k i_m} = 0 \implies \sum_{j \neq i_m} x_{i_m j} - \left(\sum_{k<i_m} x_{k i_m} + \sum_{k>i_m} x_{k i_m} \right) = 0$$

$$\implies x_{i_m j_m} - \left(\sum_{k<j_m} x_{k i_m} + \sum_{k>j_m} x_{k i_m} \right) \leq 0$$

Using this and Eq. 25, it follows that

$$0 < x_{i_m j_m} \leq \sum_{k>i_m} x_{k i_m} \implies \sum_{k>i_m} x_{k i_m} > 0.$$

Hence, there is (at least) $k^{'} > i_m$ such that $x_{k^{'} i_m} > 0$. So, $l_{k^{'} i_m} > 0 \implies \mathbf{L} \neq \mathbf{0}$. ∎

Lemma A.3 *Assume* $f(\mathbf{X}) = \sum_{i=0}^{d} \sum_{j=0}^{d} a_{ij} x_{ij}$ *and* $a_{ij} = \frac{1}{\psi(\omega)} e^{-\omega i} c(j)$ *where* $\omega \in (0, \infty)$ *and* $c(x|\boldsymbol{\lambda})$ *is a decreasing connection probability function. Consider this constraint integer programming minimization problem:*

$$\begin{aligned} \min_{\mathbf{X} \in (\mathbb{N} \cup \{0\})^{(d+1)\times(d+1)}} \quad & f(\mathbf{X}) \\ \textit{subject to:} \quad & \forall i \in \{0, \ldots, d\}, \ \sum_{j=0}^{d} x_{ij} = c_i \\ & \forall j \in \{0, \ldots, d\}, \ \sum_{i=0}^{d} x_{ij} = c_j \\ & \forall i, j \in \{0, \ldots, d\}, \ x_{ij} \geq 0 \end{aligned} \tag{26}$$

where $\forall i$, c_i *is a positive integer constant. Specifically, it is given that* $c_0 = 1$ *and* $c_d \geq 1$. *Matrix* $\mathbf{X}^*$ *is the solution to this minimization problem with* $x^*_{0d} = c_0 = 1$ *and* $x^*_{id} = 0$ *for* $i < d$.

Proof of Lemma A.3 The proof is by contradiction and its general idea is as follows. We assume that $\mathbf{X}^*$ is a feasible solution and minimizes the objective function $f(.)$ but does not satisfy $x^*_{0d} = c_0 = 1$ and $x^*_{id} = 0$ for $i < d$. Then, by slight modification of some elements in $\mathbf{X}^*$, we create matrix $\mathbf{Y}^*$ and show that it is indeed feasible solution. Afterwards, we show that $f(\mathbf{Y}^*) < f(\mathbf{X}^*)$ which contacts the minimality of $\mathbf{X}^*$.

Assume that $\mathbf{X}^*$ minimizes the objective function $f(.)$ but does not satisfy $x^*_{0d} = c_0 = 1$ and $x^*_{id} = 0$ for $i < d$. So, based on the row constraint of $\sum_{j=0}^{d} x^*_{0j} = c_0 = 1$, there should be $l < d$ such that $x^*_{0l} = 1$. Similarly, as $x^*_{0d} = 0$ and $\sum_{i=0}^{d} x^*_{id} = c_d \geq 1$, there should be $k > 0$ such that $x^*_{kd} \geq 1$. Let

$$B = \{(0, l), (0, d), (k, l), (k, d)\},$$

, we can define matrix $\mathbf{Y}^*$ as follows:

$$y^*_{ij} = \begin{cases} x^*_{ij}, & (i,j) \notin B \\ 0, & (i,j) = (0,l) \\ 1, & (i,j) = (0,d) \\ x^*_{kd} - 1, & (i,j) = (k,d) \\ x^*_{kl} + 1, & (i,j) = (k,l) \end{cases} \tag{27}$$

It is almost straightforward to see that $\mathbf{Y}^*$ satisfies all constraints stated in Eq. 26. We can first observe that $y^*_{ij} \geq 0$ and $y^*_{ij} \in \mathbb{N} \cup \{0\}$ for all $(i,j) \in B$. Moreover, $y^*_{ij} = x^*_{ij} \geq 0$ for all $(i,j) \notin B$. Since $\mathbf{Y}^*$ is different than $\mathbf{X}^*$ only in four corners of a box determined by coordinates in B, to show that $\mathbf{Y}^*$ satisfy all other constraints, it is only sufficient to demonstrate the constraints related to rows 0 and k, and columns l and d are satisfied (note that as the other rows and columns are unchanged compared to $\mathbf{X}^*$, their corresponding constraints should be already satisfied.). From definition of $\mathbf{Y}^*$ in Eq. 27, we observe that whenever 1 is added to the original x^*_{ij}, it is subtracted from another element in the same row and another element in the same column; this has guaranteed that the constraints are still met.

As $\mathbf{Y}^*$ is a feasible solution, we can write:

$$\begin{aligned} f(\mathbf{Y}^*) - f(\mathbf{X}^*) &= \sum_i \sum_j y^*_{ij} a_{ij} - \sum_i \sum_j x^*_{ij} a_{ij} = \sum_{(i,j)\in B} a_{ij}(y^*_{ij} - x^*_{ij}) \\ &= a_{0d} - a_{0l} - a_{kd} + a_{kl} \\ &= \frac{1}{\psi(\omega)} \left[(c(d) - c(l)) - \left(e^{-\omega k} c(d) - e^{-\omega k} c(l) \right) \right] \\ &= \frac{\delta}{\psi(\omega)} \left(1 - e^{-\omega k} \right) (c(d) - c(l)) \end{aligned}$$

Since $d > l$ and $c(.)$ is decreasing function, $c(d) - c(l) < 0$. Also, since $0 < k$ and $\omega \in (0, \infty)$, we have $1 - e^{-\omega k} > 0$. Hence,

$$\frac{1}{\psi(\omega)} \left(1 - e^{-\omega k} \right) (c(d) - c(l)) < 0 \implies f(\mathbf{Y}^*) > f(\mathbf{X}^*),$$

which contradicts the optimality of $\mathbf{X}^*$. So matrix $\mathbf{X}^*$ must satisfy $x^*_{0d} = c_0 = 1$ and $x^*_{id} = 0$ for $i < d$. ∎

Proof of Proposition 3.4 The proof is almost trivial and follows for Thm. 3.3 and Eq. 3:

$$\begin{aligned} &\mathcal{D}(\sigma_M, \boldsymbol{\theta}) \leq \mathcal{D}(r, \boldsymbol{\theta}) \leq \mathcal{D}(\sigma, \boldsymbol{\theta}) \\ &\implies \sum_{r \in \Omega(\mathcal{A})} \rho(r|\boldsymbol{\theta}) \mathcal{D}(\sigma_M, \boldsymbol{\theta}) \leq \sum_{r \in \Omega(\mathcal{A})} \rho(r|\boldsymbol{\theta}) \mathcal{D}(r, \boldsymbol{\theta}) \leq \sum_{r \in \Omega(\mathcal{A})} \rho(r|\boldsymbol{\theta}) \mathcal{D}(\sigma, \boldsymbol{\theta}) \\ &\implies \mathcal{D}(\sigma_M, \boldsymbol{\theta}) \sum_{r \in \Omega(\mathcal{A})} \rho(r|\boldsymbol{\theta}) \leq \mathcal{E}(\boldsymbol{\theta}) \leq \mathcal{D}(\sigma, \boldsymbol{\theta}) \sum_{r \in \Omega(\mathcal{A})} \rho(r|\boldsymbol{\theta}) \\ &\implies \mathcal{D}(\sigma_M, \boldsymbol{\theta}) \leq \mathcal{E}(\boldsymbol{\theta}) \leq \mathcal{D}(\sigma, \boldsymbol{\theta}) \quad ∎ \end{aligned}$$

B Rational Behind Approximations

We here explain the main idea and rational behind our approximation methods.

Approximation to Network Diameter. We know that two nodes have the lowest chance of connectivity when their rankings have the maximum possible distance $d_M(m)$. In other words, the probability that any pair of nodes get acquainted is at least $c(d_M(m), \boldsymbol{\lambda})$. We also note that by increasing the edge probability, the diameter (or the average shortest path length) does not increase (actually, most often it decreases). So one can bound the diameter $D(\theta)$ (or the average shortest path length $\langle l(\theta) \rangle$) by the diameter $D^{'}(n,p)$ (or the average shortest path length $\langle l^{'}(n,p) \rangle$) of the random graph $G(n,p)$ with fixed $p = c(d_M(m), \boldsymbol{\lambda})$. So $D(\boldsymbol{\theta}) \leq D^{'}(n,p)$ and $\langle l(\boldsymbol{\theta}) \rangle \leq \langle l^{'}(n,p) \rangle$ with $p = c(d_M(m), \boldsymbol{\lambda})$.

To give more clear picture, one can assume that the network formation – under ranking network model– runs in two phases. First any two nodes get connected to each other with fixed probability of $c(d_M(m), \boldsymbol{\lambda})$ (similar to random graph model) which results in diameter, say, $D^{'}(n,p)$ and average shortest path length, say, $\langle l^{'}(n,p) \rangle$. Then, any pair of nodes with ranking r and $r^{'}$ get connected with the probability $c(d(r, r^{'}), \boldsymbol{\lambda}) - c(d_M(m), \boldsymbol{\lambda})$. The later phase results in final diameter $D(\theta)$ and average shortest path length $\langle l(\theta) \rangle$. Obviously adding more edge in the second phase can not

increase the diameter and the average shortest path. So $D(\boldsymbol{\theta}) \leq D^{'}(n,p)$ and $\langle l(\theta)\rangle \leq \langle l^{'}(n,p)\rangle$ with $p = c(d_M(m), \boldsymbol{\lambda})$.

One can approximate the diameter $D^{'}(n,p)$ and the average shortest path length $\langle l^{'}(n,p)\rangle$ in $G(n,p)$ model by

$$D^{'}(n,p) \approx \left\lceil \frac{\log(n)}{\log(n-1)p} \right\rceil \ and \ \langle l^{'}(n,p)\rangle \approx \frac{\log(n)}{\log(n-1)p}.$$

Setting $p = c(d_M(m), \boldsymbol{\lambda})$ and using the inequalities $D(\theta) \leq D^{'}(n,p)$ and $\langle l(\theta)\rangle \leq \langle l^{'}(n,p)\rangle$, we have shown that

$$D(\boldsymbol{\theta}) \leq \left\lceil \frac{\log(n)}{\log(n-1) + \log c(d_M(m)|\boldsymbol{\lambda})} \right\rceil \ and \ \langle l(\boldsymbol{\theta})\rangle \leq \frac{\log(n)}{\log(n-1) + \log c(d_M(m)|\boldsymbol{\lambda})}.$$

$\tilde{\tilde{E}}$ **Approximation.** Let $p(d(r,\sigma) = i|\omega,\sigma)$ denote the probability that the drawn ranking r from the distance-based ranking model $\rho(r|\omega,\sigma)$ has distance i to the reference ranking σ. We can note that for $i \in \{0, \cdots, d_M(m)\}$,

$$p(d(r,\sigma) = i|\omega,\sigma) = n_i \frac{e^{-i\omega}}{\psi(\omega)},$$

where n_i represents the number of rankings with distance of i to the reference ranking σ. One can approximate $p(d(r,\sigma) = i|\omega,\sigma)$ by a binomial distribution

$$B\left(i|d_M(m), \frac{e^{-\omega}}{1+e^{-\omega}}\right) = \ binomd_M(m)i \frac{e^{-i\omega}}{(1+e^{-\omega})^{d_M(m)}}$$

where n_i, and $\psi(\omega)$ are approximated by $\binom{d_M(m)}{i}$ and $(1+e^{-\omega})^{d_M(m)}$ respectively. Using this approximation, the convexity of function of $c(.|\boldsymbol{\lambda})$, and Equations 13, 10, and 12, we can write

$$\begin{aligned}
\tilde{\mathcal{D}}(x,\boldsymbol{\theta}) &= \left(1 - \frac{x}{d_M(m)}\right)\mathcal{D}(\sigma,\boldsymbol{\theta}) + \left(\frac{x}{d_M(m)}\right)\mathcal{D}(\sigma_M,\boldsymbol{\theta}) \\
&\approx \left(1 - \frac{x}{d_M(m)}\right)\sum_{i=0}^{d_M(m)} \binom{d_M(m)}{i} \frac{e^{-i\omega}}{(1+e^{-\omega})^{d_M(m)}} c(i|\boldsymbol{\lambda}) \\
&+ \left(\frac{x}{d_M(m)}\right)\sum_{i=0}^{d_M(m)} \binom{d_M(m)}{i} \frac{e^{-i\omega}}{(1+e^{-\omega})^{d_M(m)}} c(d_M(m) - i|\boldsymbol{\lambda}) \\
&= \left(1 - \frac{x}{d_M(m)}\right) \underset{x\sim B\left(d_M(m), \frac{e^{-\omega}}{1+e^{-\omega}}\right)}{\mathbb{E}\left[c(x,\boldsymbol{\lambda})\right]} + \frac{x}{d_M(m)} \underset{x\sim B\left(d_M(m), \frac{e^{-\omega}}{1+e^{-\omega}}\right)}{\left[c(d_M(m) - x,\boldsymbol{\lambda})\right]} \\
&\leq \left(1 - \frac{x}{d_M(m)}\right) c(\mathbb{E}[x],\boldsymbol{\lambda}) + \left(\frac{x}{d_M(m)}\right) c(d_M(m) - \mathbb{E}[x],\boldsymbol{\lambda})
\end{aligned}$$

We let $\tilde{\tilde{\mathcal{D}}} = \left(1 - \frac{x}{d_M(m)}\right) c(\mathbb{E}[x],\boldsymbol{\lambda}) + \left(\frac{x}{d_M(m)}\right) c(d_M(m) - \mathbb{E}[x],\boldsymbol{\lambda})$ Now, as $\underset{x\sim B\left(d_M(m), \frac{e^{-\omega}}{1+e^{-\omega}}\right)}{\mathbb{E}\left[x\right]} = \frac{d_M(m)e^{-\omega}}{1+e^{-\omega}}$, we have

$$\tilde{\tilde{\mathcal{D}}}(x,\boldsymbol{\theta}) = \left(1 - \frac{x}{d_M(m)}\right) c\left(\frac{d_M(m)e^{-\omega}}{1+e^{-\omega}}\right) + \frac{x}{d_M(m)} c\left(\frac{d_M(m)}{1+e^{-\omega}}\right).$$

Using this approximation, one can approximate the edge density by

$$\begin{aligned}
\mathcal{E}(\tilde{\tilde{\boldsymbol{\theta}}}) &= \underset{x\sim B\left(d_M(m), \frac{e^{-\omega}}{1+e^{-\omega}}\right)}{\mathbb{E}\left[\tilde{\tilde{\mathcal{D}}}(x,\boldsymbol{\theta})\right]} \\
&= \frac{1}{1+e^{-\omega}} c\left(\frac{d_M(m)e^{-\omega}}{1+e^{-\omega}}\right) + \frac{e^{-\omega}}{1+e^{-\omega}} c\left(\frac{d_M(m)}{1+e^{-\omega}}\right)
\end{aligned}$$

C Computing the number of rankings with distance k for Kendall-τ distance

We here explain our dynamic programming approach for calculating the number of rankings with Kendall-τ distance k from an arbitrary fixed reference ranking. We let $T(j,k)$ denote the number of rankings with Kendall tau distance $k \in [0, \binom{j}{2}]$ from a reference ranking with j alternatives. We note that $T(j,0) = 1$ for any $j \geq 1$ and $T(j,k) = 0$ for any $k > \binom{j}{2}$. We focus on the fact that Kendall tau distance between two permutations measures the minimum number of swaps or inversions for converting one permutation to the other permutation. Considering the insertion model, we also observe that the j^{th} alternative can be inserted in j possible positions, thus causing 0 to $j-1$ inversions or swaps (0 to $j-1$ distance). This implies that when $k \geq j-1$, the j^{th} alternative can be part of either $0, 1, \ldots, or\ j-1$ inversions while the other preceding $j-1$ alternatives involve in $k-0, k-1, \ldots, or\ k-j+1$ inversions. Similarly, when $k < j-1$, the j^{th} alternative can be part of either $0, 1, \ldots, or\ k$ inversions while the other preceding $j-1$ alternatives involve in $k, k-1, \ldots, or\ 0$ inversions. Hence, we can write this recursive equation:

$$T(j,k) = \sum_{i=0}^{\min(k,j-1)} T(j-1, k-i) \tag{28}$$

Since $k \in [0, \binom{m}{2}]$ and $j \in [1,m]$, T can be viewed as a $m \times \frac{m(m-1)}{2}$ matrix. For calculating each element of this matrix, at most $m-1$ summation operation is required. Thus, the running time of our dynamic approach is $O(m^4)$. However, this time complexity can be improved by slight change of our recursive formula. From Eq. 28, we note that for $k \leq j-1$, $T(j,k) - T(j,k-1) = T(j-1,k)$ and for $k > j-1$, $T(j,k) - T(j,k-1) = T(j-1,k) - T(j-1,j-k)$. Hence, the recursive formula in Eq. 28 is equivalent to:

$$T(j,k) = \begin{cases} T(j,k-1) + T(j-1,k), & k \leq j-1 \\ T(j,k-1) + T(j-1,k) - T(j-1,k-j), & k > j-1 \end{cases} \tag{29}$$

Using this recursion, each element in matrix T can be calculated in $O(1)$, thus yielding in overall time complexity of $O(m^3)$. Note that, in this manuscript, we are interested in those values located in row m of matrix T (i.e., $T(m,k)$ for $0 \leq k \leq \binom{m}{2}$), thus using n_k notation instead of $T(m,d)$ so far and on.