GENERATING SIMILAR GRAPHS FROM SPHERICAL FEATURES

Dalton Lunga and Sergey Kirshner
{dlunga,skirshne}@purdue.edu

Summary

We propose a novel model for generating graphs similar to a given example graph. Unlike standard approaches that compute features of graphs in Euclidean space, our approach obtains features on a surface of a hypersphere. We then utilize a von Mises-Fisher distribution, an exponential family distribution on the surface of a hypersphere, to define a model over possible feature values. While our approach bears similarity to a popular exponential random graph model (ERGM), unlike ERGMs, it does not suffer from degeneracy, a situation when a significant probability mass is placed on unrealistic graphs. We propose a parameter estimation approach for our model, and a procedure for drawing samples from the distribution. We evaluate the performance of our approach both on the small domain of all 8-node graphs as well as larger real-world social networks.

Contents

List of Figures

List of Tables

1   Introduction

Increasingly, many domains produce data sets containing relationships that are conveniently represented by networks, e.g., systems sciences (the Internet), bioinformatics (protein interactions), social domains (social networks). As researchers in these areas are developing models and tools to analyze the properties of networks, they are hampered by few samples available to evaluate their approaches. This gives rise to a problem of generating more network samples that can be viewed as drawn from the same population as the given network.

While there are a number of possible approaches to this problem, perhaps the most well-studied model is the exponential random graph models (ERGMs, or in social network literature, $p^\star$), an exponential family class of models, matching the statistics over the set of possible networks to the statistics of the network in question [e.g., 17]. This and similar approaches have a long history as they generalize the p_1 [7] and Markov random graph [4] models first developed in the social network literature. While such approaches are intuitive and have nice properties, they also suffer from the issue of *degeneracy* [5, 16], which is manifested in the instability of parameter estimation, and in placing most probability mass of resulting distributions on unrealistic graphs (e.g., empty graph or complete graphs). As a result, these approaches are not suitable for the purpose of generating graphs similar to the given one.

This paper contains two contributions. First, we first zero in on the issue of degeneracy and discover that its cause is related to the geometry of the set of feature vectors and the number of graphs mapped to each feature vector (thought of as feature vector weights). By augmenting feature vectors with logarithms of their corresponding weights, we show that only graphs with such augmented feature vectors on the relative boundary of the resulting extended convex hull can become modes of any exponential random graph model, explaining why unrealistic graphs (which are on the relative boundary) often get large probability masses. Second, using the insight of the observation above, we propose a novel random graph model which is based on embedding the features of graphs onto a surface of a hypersphere. Since a spherical surface is a relative boundary of the sphere's convex hull, all of the feature vectors would then belong to the relative boundary of the convex hull and could potentially serve as modes of the corresponding distributions. This in turn helps to avoid the degeneracy issues which plague ERGMs.

Our proposed approach makes use of *spherical* features obtained by embedding possible graphs onto a surface of a sphere [18], and then approximating the distribution of the resulting spherical feature space with a von Mises-Fisher distribution. Since the space of all possible graphs is too large to consider for embedding, we consider determining the embedding function based only on the neighborhood around the given graph thus resulting in a locally spherical embedding of the set of graphs. The main benefit of our approach is that it fixes the issue of degeneracy, with the mode of the distribution over the spherical feature vector coinciding with the features

of the given graph. An additional advantage of our approach is that its parameter estimation procedure does not require cumbersome maximum entropy approaches used with ERGMs.

We start by revisiting the ERGM model and presenting insights on why this model often fails to generate realistic graphs (Section 2). We then propose our alternative approach, exponential locally spherical random graph model (`ELSRGM`, Section 3) and evaluate it on both synthetic and realistic graphs while comparing it to ERGMs (Section 4). We conclude the paper (Section 5) with ideas for future work.

2 ERGMs

Overall, we are interested in probabilistic approaches for generating graphs similar to the given one. We consider the case of simple (unweighted, no self-loops) undirected graphs $G = (V, E) \in \mathcal{G}_n$.[1] Where $n = |V|$ is the number of vertices; $E \in \{0,1\}^{n \times n}$ is a symmetric binary adjacency matrix with zeros on its diagonal, with $e_{ij} = 1$ iff there is an edge from v_i to v_j, and $e_{ij} = 0$ otherwise, where $v_i \in V$ and $e_{ij} \in E$. There are $|\mathcal{G}_n| = 2^{\binom{n}{2}}$ possible labeled graphs with n vertices, a finite but often prohibitively large number even for fairly small n.

We first consider the well-studied exponential random graph model (ERGM, also known as $p^\star$) as a starting point for our approach.

2.1 ERGM Definition

In the area of social network analysis, scientists are often interested in specific features of networks, and some of the state-of-the-art models explicitly use them to define functions of network sub-structures. We will denote the vector of these functions by $\boldsymbol{f} : \mathcal{G}_n \to \mathbb{R}^d$. Among the examples of such features used by social scientists, we have the number of edges, the number of triangles, the number of k-stars, etc.:

$$
\begin{aligned}
f_{edge}(G) &= \sum \sum_{1 \leq < i < j \leq n} e_{ij}; \\
f_\triangle(G) &= \sum \sum \sum_{1 \leq i < j < k \leq n} e_{ij} e_{ik} e_{jk}; \\
f_{k\star}(G) &= \sum_{i=1}^{n} \sum_{\substack{1 \leq i_1 < \cdots \leq i_k \\ i_1 \neq i, \ldots, i_k \neq i}} \cdots \sum \prod_{j=1}^{k} e_{ii_j}.
\end{aligned}
\tag{1}
$$

The subgraph patterns corresponding to features in equation (1) are shown in Figure 1.

A commonly used probabilistic model over $\mathcal{G}_n$ is an exponential family model that uses the expectation of $\boldsymbol{f}$ as a vector of sufficient statistics. The distribution over $\mathcal{G}_n$

---

[1] Our approach extends to directed graphs as well.

Figure 1: Simple typical subgraph configurations for undirected graphs. From *left* to *right*: `edge`, `triangle`, `two-star` and `three-star` configurations.

can be parameterized in the form

$$P\left(G|\boldsymbol{\theta}\right) = \frac{\exp\left\langle\boldsymbol{\theta}, \boldsymbol{f}\left(G\right)\right\rangle}{Z\left(\boldsymbol{\theta}\right)}, \ G \in \mathcal{G}_n, \tag{2}$$

where $\boldsymbol{\theta} \in \mathbb{R}^d$ is the vector of natural model parameters, $\boldsymbol{f}\left(G\right)$ is a vector of features of G, and

$$Z\left(\boldsymbol{\theta}\right) = \sum_{G \in \mathcal{G}_n} \exp\left\langle\boldsymbol{\theta}, \boldsymbol{f}\left(G\right)\right\rangle$$

is the partition function.

Let $\mathcal{P} = \{\boldsymbol{f}\left(G\right) : G \in \mathcal{G}_n\}$ denote the set of possible feature-vector values. Under ERGMs, the space $\mathcal{G}_n$ of possible graphs is coarsened into a much smaller set $\mathcal{P}$ of possible features. The distribution in (2) can also be viewed as a distribution over $\mathcal{P}$ with all of the graphs mapped to the same value of $\boldsymbol{f}$ assigned the same probability mass. Let $w\left(\boldsymbol{x}\right) = |\{G \in \mathcal{G}_n : \boldsymbol{f}\left(G\right) = \boldsymbol{x}\}|$ be the number of graphs corresponding to a feature value $\boldsymbol{x} \in \mathcal{P}$. (We will refer to $w\left(\boldsymbol{x}\right)$'s as *weights*.) The distribution $P\left(G|\boldsymbol{\theta}\right)$ can also be extended to the set $\mathcal{P}$,

$$\begin{aligned}
P\left(\boldsymbol{x}|\boldsymbol{\theta}\right) &= \frac{1}{Z\left(\boldsymbol{\theta}\right)} w\left(\boldsymbol{x}\right) \exp\left\langle\boldsymbol{\theta}, \boldsymbol{x}\right\rangle, \ \boldsymbol{x} \in \mathcal{P}, \\
Z\left(\boldsymbol{\theta}\right) &= \sum_{\boldsymbol{x} \in \mathcal{P}} w\left(\boldsymbol{x}\right) \exp\left\langle\boldsymbol{\theta}, \boldsymbol{x}\right\rangle,
\end{aligned} \tag{3}$$

also an exponential family distribution. We let $\mathcal{C} \subseteq \mathbb{R}^d$ denote a convex hull of $\mathcal{P}$. Since $\mathcal{G}_n$ is finite, so would be $\mathcal{P}$, and $\mathcal{C}$ would be a polytope in $\mathbb{R}^d$. Using a well-known result from the theory of exponential families [1], if $\boldsymbol{x}^\star \in \mathbb{R}^d$ is a vector of the sufficient statistics (e.g., for a single example graph $G^\star$, $\boldsymbol{x}^\star = \boldsymbol{f}\left(G^\star\right)$), a maximum likelihood estimate (MLE) $\hat{\boldsymbol{\theta}}\left(\boldsymbol{x}^\star\right)$ satisfies

$$E_{P\left(G|\hat{\boldsymbol{\theta}}(\boldsymbol{x}^\star)\right)}\boldsymbol{f}\left(G\right) = E_{P\left(\boldsymbol{x}|\hat{\boldsymbol{\theta}}(\boldsymbol{x}^\star)\right)}\boldsymbol{x} = \boldsymbol{x}^\star, \tag{4}$$

and it exists, and the corresponding distribution $P\left(\cdot|\hat{\boldsymbol{\theta}}\left(\boldsymbol{x}^\star\right)\right)$ is unique as long as $\boldsymbol{x}^\star \in \mathrm{rint}\left(\mathcal{C}\right)$, the relative interior of the convex hull of the set of possible feature vectors $\mathcal{P}$. In essence, ERGMs are designed to preserve mean features of the observed graph, a very intuitive and often desirable property. However, it is also well-known

that the estimation of $\hat{\boldsymbol{\theta}}\left(\boldsymbol{x}^{\star}\right)$ is an extremely cumbersome task, complicated by the fact that the exact calculation of $Z\left(\boldsymbol{\theta}\right)$ is intractable, and approximate approaches (e.g., pseudo-likelihood, MCMC) are employed instead [9].

2.2 Degeneracy in ERGMs

ERGMs often suffer from the problem of *degeneracy* [5]. There are two types of degeneracy usually considered. The first type occurs when the MLE estimate $\hat{\boldsymbol{\theta}}\left(\boldsymbol{x}^{\star}\right)$ either does not exist, or the MLE estimation procedure does not converge due to numerical instabilities [5, 16].

The second type of degeneracy happens when $\hat{\boldsymbol{\theta}}\left(\boldsymbol{x}^{\star}\right)$ can be reliably estimated, but the resulting probability distribution places significant probability mass (or virtually *all* of probability mass) on unrealistic graphs, e.g., empty or complete graphs. This type of degeneracy can be considered from another viewpoint; that is the mode of ERGM corresponding to the $\hat{\boldsymbol{\theta}}\left(\boldsymbol{x}^{\star}\right)$ may be placed on $\boldsymbol{x} \in \mathcal{P}$ very different from $\boldsymbol{x}^{\star}$. This is an undesirable property as there is little justification for placing large probability mass over a region away from observed feature vectors while placing little mass over the observed example.

For an illustration, consider the set $\mathcal{G}_8$ (the set of all possible simple undirected graphs with 8 nodes) with feature vectors $\boldsymbol{f}\left(G\right) = \left(f_{edge}\left(G\right), f_{\triangle}\left(G\right)\right)$. In this case, Figure 2 displays (left plot) the support space $\mathcal{P}$ consists of 228-(*edge,triangle*) pair-statistics for all possible 8-node undirected graphs. The right plot of Figure 2, displays a diameter distribution highlighting that graphs with different topologies map to same *edge-triangular* feature pairs. The diameter was computed by observing that each feature-pair could be viewed as a cell of graphs with same feature count. This can be accomplished by first computing a perturbation graph whose nodes are all non-isomorphic graphs of 8-nodes. A graph edit distance is applied on the set $\mathcal{G}_8$ containing 12346-non isomorphic graphs (computed using `nauty` [13]). To compute the diameter for each cell given the perturbation graph, one only has to identify the graphs mapping to that cell and extract the maximum number of edges for any pair of graphs with their feature counts mapping to that cell. Table 1 shows the complexity of small sized graph spaces.

Table 1: Complexity of graph spaces for fixed number of nodes.

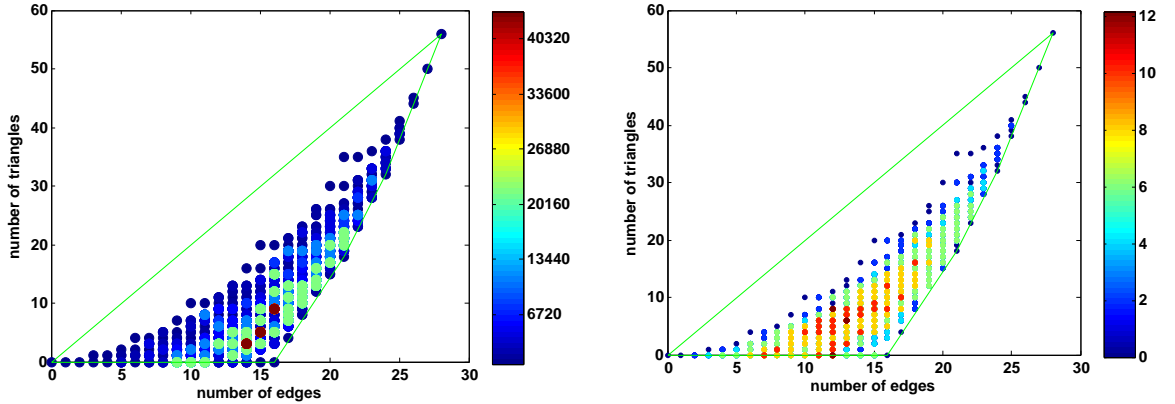| n	number of edges	all graphs are $2^{\binom{n}{2}}$	non-isomorphic
7	21	2,097,152	1,044
8	28	268,435,456	12,346
9	36	68,719,476,736	274,668
10	45	35,184,372,088,832	12,005,168
11	55	3,602,879,701,896,397	1,018,997,864
12	66	7,378,697,629,483,821,000	165,091,172,592

Figure 2: *left:* Distribution of 8-node graphs. *right:* Distribution of cell diameter; where we consider each feature-pair as a cell and compute the graph edit distance between each pair of graphs with feature counts mapping to that cell (feature-pair).

Figure 4 shows the support $\mathcal{P}$ for the distribution over feature-vectors (all circles), and its convex hull $\mathcal{C}$ (boundary in green). Only a small subset of the example feature pairs $\boldsymbol{x}^{\star} = \boldsymbol{f}(G^{\star})$ result in distributions over $\mathcal{P}$ with modes coinciding the example feature vector $\boldsymbol{x}^{\star}$ (red discs with black borders). ERGMs estimated from other feature vectors will thus generate graphs with different features from the example graph as shown in Figure 3.

It is this behavior that we are most concerned with and are trying to address in this paper, and from now on when we mention degeneracy, unless otherwise noted, we mean the second type of degeneracy.

2.2.1   Approaches to Fixing Degeneracy in ERGMs

Recent approaches in literature have focused on a more flexible specification of the model in equation (3). This has been achieved by a mixed set of feature statistics which includes the node attributes of a given graph [6]. Instead of using only the structural properties (*edges, triangles etc.*) of the graph, other attributes (e.g., *gender,race,age,etc*) are included as covariates for the model. This approach, however, does not address degeneracy in general as one has to know what set of features and attributes to choose in attempting to minimize degeneracy [6]. Such a task demands domain knowledge to accurately specify a reasonable model. Another approach makes use of the geometrically weighted edgewise shared partner, the geometrically weighted dyadic shared partner, and the geometrically weighted degree network statistics as new statistics for the model in (2). These specifications when parameterized have been shown to lead to curved exponential families [9]. However, the difficulty with this approach is not only in the parameter estimation of the resulting curved exponential model, but also on having to avoid other graph features that may be dependent to these specifications as that will result in degeneracy [8].
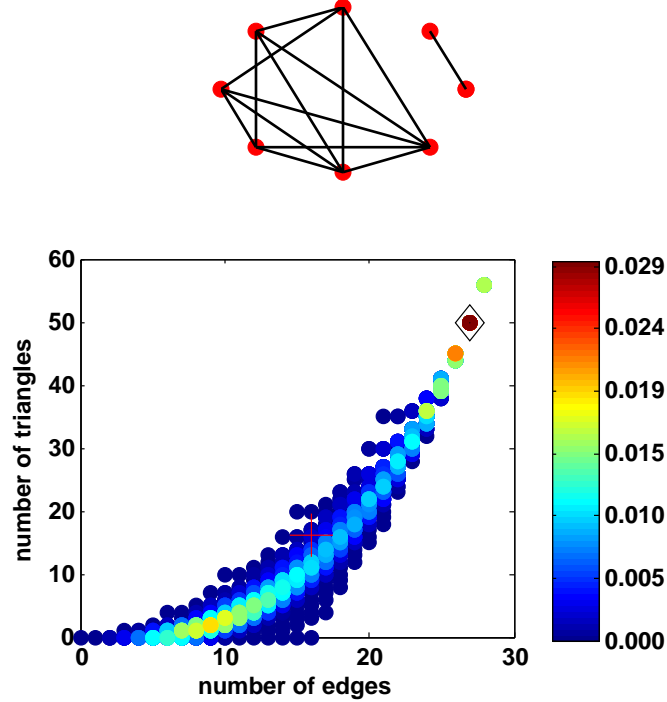
Figure 3: A degenerate ERGM specified by edge-triangle pair for an 8-node graph(*top figure*). Colorcoded is the pmf over the edge-triangle space for an estimated MLE $\theta_{MLE} = (-0.992, 0.617)$. "+" indicates the observed feature and its mean, while the "⋄" shape indicates the ERGM mode.

2.3 Why Are Unrealistic Graphs Likely under ERGMs?

It may not be entirely surprising that the probability mass over support $\mathcal{P}$ does not concentrate around the expected value of P as the mode and the mean of the exponential family distributions do not always match. This does not however imply that the mode is placed on features corresponding to unrealistic graphs. To investigate the degeneracy further, we first introduce the following result from convex analysis.

Theorem 1. *Suppose $\mathcal{C} \subset \mathbb{R}^d$ is a full-dimensional bounded convex polytope with a finite set of vertices $\mathcal{V} = \left\{ \boldsymbol{x}^1, \ldots, \boldsymbol{x}^L \right\}$. Then for any $\boldsymbol{\theta} \in \mathbb{R}^d \setminus \{\boldsymbol{0}\}$,*

$$\mathcal{M}_{\boldsymbol{\theta}} = \operatorname*{argmax}_{\boldsymbol{x} \in \mathcal{C}} \langle \boldsymbol{\theta}, \boldsymbol{x} \rangle \subset rbd\left(\mathcal{C}\right),$$

where by $rbd\left(\mathcal{C}\right)$ we denote the relative boundary of the convex hull $\mathcal{C}$.

Proof. Since $\mathcal{C}$ is a convex hull of $\mathcal{V}$, for all $\boldsymbol{x} \in \mathcal{C}$

$$\boldsymbol{x} = \sum_{l=1}^{L} \lambda_l \boldsymbol{x}^l, \text{ for some } \lambda_1, \ldots, \lambda_L \geq 0, \ \sum_{l=1}^{L} \lambda_l = 1. \tag{5}$$

Figure 4: Solid points show mode placement on the edge-triangle feature pairs that form the vertices of the extended hull as a result of Theorem 1. Overlaid on solid points are modes obtained from using the estimated $\hat{\theta}_{mle}$ of equation (3).

For any $\boldsymbol{\theta} \in \mathbb{R}^d \setminus \{\mathbf{0}\}$, let $\mathcal{V}_{\boldsymbol{\theta}} = \operatorname{argmax}_{\boldsymbol{x} \in \mathcal{V}} \langle \boldsymbol{\theta}, \boldsymbol{x} \rangle$, and let $a_{\boldsymbol{\theta}} = \max_{\boldsymbo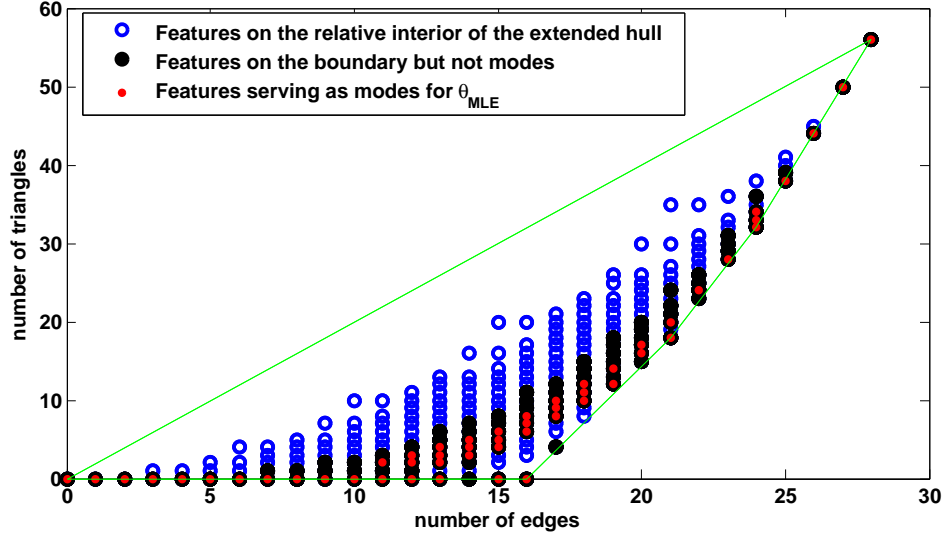l{x} \in \mathcal{V}} \langle \boldsymbol{\theta}, \boldsymbol{x} \rangle$. Then $\forall \; \boldsymbol{x} \in \mathcal{C}$, $\forall \boldsymbol{x}_\theta \in \mathcal{V}_{\boldsymbol{\theta}}$

$$\langle \boldsymbol{\theta}, \boldsymbol{x} \rangle = \left\langle \boldsymbol{\theta}, \sum_{l=1}^{L} \lambda_l \boldsymbol{x}^l \right\rangle = \sum_{l=1}^{L} \lambda_l \langle \boldsymbol{\theta}, \boldsymbol{x}^l \rangle \le \sum_{l=1}^{L} \lambda_l \langle \boldsymbol{\theta}, \boldsymbol{x}_{\boldsymbol{\theta}} \rangle$$
$$= \langle \boldsymbol{\theta}, \boldsymbol{x}_{\boldsymbol{\theta}} \rangle = a_{\boldsymbol{\theta}}.$$

Then $\left\{ x \in \mathbb{R}^d : \; \langle \boldsymbol{\theta}, \boldsymbol{x} \rangle \le a_{\boldsymbol{\theta}} \right\}$ is a supporting hyperspace, and $\mathcal{H}\left(a_{\boldsymbol{\theta}}\right) = \left\{ x \in \mathbb{R}^d : \; \langle \boldsymbol{\theta}, \boldsymbol{x} \rangle = a_{\boldsymbol{\theta}} \right\}$ is a supporting $d-1$ dimensional hyperplane for $\mathcal{C}$. Since $\mathcal{C}$ is a full-dimensional convex polytope, $\mathcal{C} \not\subset \mathcal{H}\left(a_{\boldsymbol{\theta}}\right)$, and therefore, $\mathcal{H}\left(a_{\boldsymbol{\theta}}\right)$ is a proper supporting hyperplane. Thus $\operatorname{rint}\left(\mathcal{C}\right) \cap \mathcal{H}\left(a_{\boldsymbol{\theta}}\right) = \emptyset$ (e.g., [2], Thm. 4.1). Observing that $\mathcal{M}_{\boldsymbol{\theta}} \subset \mathcal{H}\left(a_{\boldsymbol{\theta}}\right)$ completes the proof. $\qquad \square$

We now consider the formulation in (3), from which after taking the log-likelihood we observe that

$$\operatorname*{argmax}_{\boldsymbol{x} \in \mathcal{P}} \ln P\left(\boldsymbol{x} | \boldsymbol{\theta}\right) = \operatorname*{argmax}_{\boldsymbol{x} \in \mathcal{P}} \left\{ \langle \boldsymbol{\theta}, \boldsymbol{x} \rangle + \ln w\left(\boldsymbol{x}\right) \right\}. \tag{6}$$

We extend $\mathcal{P}$ to include the weight w. Let $\bar{\mathcal{P}} = \left\{ \left(\boldsymbol{x}, \ln w\left(\boldsymbol{x}\right)\right) : \boldsymbol{x} \in \mathcal{P} \right\}$ be the extended set of features, with $\bar{\mathcal{C}}$ being the resulting extended convex hull for $\bar{\mathcal{P}}$ and $\bar{\mathcal{V}}$ the set of vertices for $\bar{\mathcal{C}}$. Then

$$\langle \boldsymbol{\theta}, \boldsymbol{x} \rangle + \ln w\left(\boldsymbol{x}\right) = \langle \left(\boldsymbol{\theta}, 1\right), \left(\boldsymbol{x}, \ln w\left(\boldsymbol{x}\right)\right) \rangle.$$
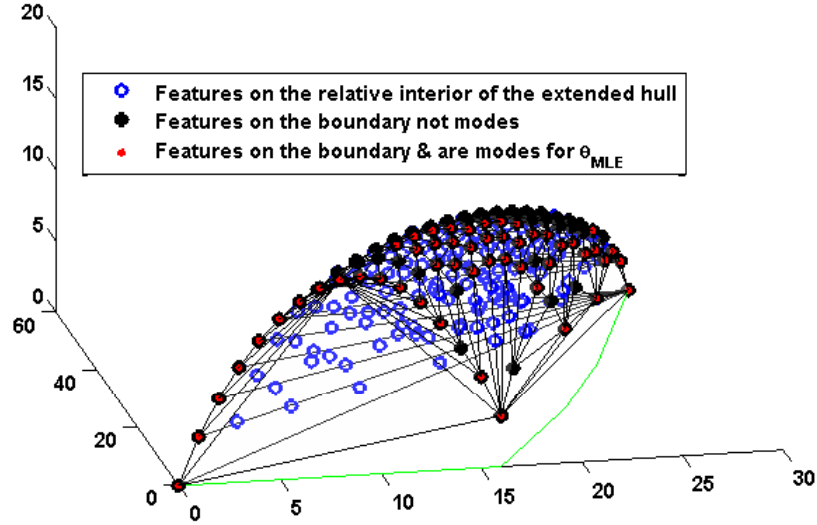
Figure 5: View of mode placement on facets of the extended 3D convex hull. Each mode forms a vertex of a given triangle (denoted by black lines) for a facet of the convex hull.

By Theorem 1, only the points on the boundary rbd $(\bar{\mathcal{C}})$ can maximize (6). Thus to find the set of possible modes of (3), we can restrict our attention only to $\mathcal{M} = \left\{ \boldsymbol{x} : (\boldsymbol{x}, \ln w (\boldsymbol{x})) \in \bar{\mathcal{V}} \right\}$.

Consider an illustration in Figure 5 for the case of all 8-node graphs. Only points on the boundary of the polytope $\bar{\mathcal{C}}$ can *potentially* be modes of any ERGM specified on the feature set in Figure 4 (points on the boundary are denoted by solid black discs and solid red discs with black borders on both plots). Figure 4 further reveals that only a small number of points on the relative boundary of $\bar{\mathcal{C}}$ correspond to *actual observed* modes for equation 3 (ERGMs with parameters $\hat{\boldsymbol{\theta}} (\boldsymbol{x}^\star)$ from (4)).[2] There are two possible reasons for this occurrence. One, some of the points in $\bar{\mathcal{P}}$ correspond to $\bar{\boldsymbol{\theta}} = (\boldsymbol{\theta}, t)$ with $t \neq 1$ (where $t \in \mathbb{R}$, is a scalar augmenting the parameter vector), so not all of $\mathcal{M}$ may be the modes of (3). Two, the MLE solution $\hat{\boldsymbol{\theta}} (\boldsymbol{x}^\star)$ may lie outside of the cone [16]

$$\left\{ \boldsymbol{\theta} \in \mathbb{R}^d : \ \boldsymbol{x}^\star \in \underset{\boldsymbol{x} \in \mathcal{C}}{\operatorname{argmax}} \langle \boldsymbol{\theta}, \boldsymbol{x} \rangle \right\}$$

of parameters for which $\boldsymbol{x}^\star$ is the mode.[3]

The above observations also explain why considering a curved exponential model

---

[2] The solid red discs in Figure 4 handles the case where modes of the maximum likelihood distribution are not unique.

[3] Often, feature vectors corresponding to the empty and to the complete graphs have particularly large cones, with many $\hat{\boldsymbol{\theta}} (\boldsymbol{x}^\star)$ values falling within these cones, where $\boldsymbol{x}^\star$ is the feature maximizing (6). This explains why many ERGMs place large probability masses on these degenerate graphs.

[8]

$$P\left(\boldsymbol{x}|\boldsymbol{\theta}\right) = \frac{1}{Z\left(\boldsymbol{\theta}\right)} w\left(\boldsymbol{x}\right) \exp \left\langle g\left(\boldsymbol{\theta}\right), \boldsymbol{x} \right\rangle, \ \boldsymbol{x} \in \mathcal{P}$$

does not rectify the issue of degeneracy as curving the space of parameters does not change the geometry of the feature space.[4]

Finally, we note that this type of degeneracy is not restricted to distributions over statistics of finite graphs. Similar issues can in general arise with exponential family distributions on a bounded support owing to the fact that the exponential family distributions are designed to match the mean statistics and not to concentrate the probability mass around the mode. Consider an illustration in Figure 6 for a 20×20 grid of uniformly spaced points; fitted is the exponential family model $p\left(\boldsymbol{x}|\boldsymbol{\theta}\right) \propto \exp \left\langle \boldsymbol{\theta}, \boldsymbol{x} \right\rangle$. As evident from the plots, the estimated model exhibit the degeneracy issue described above.
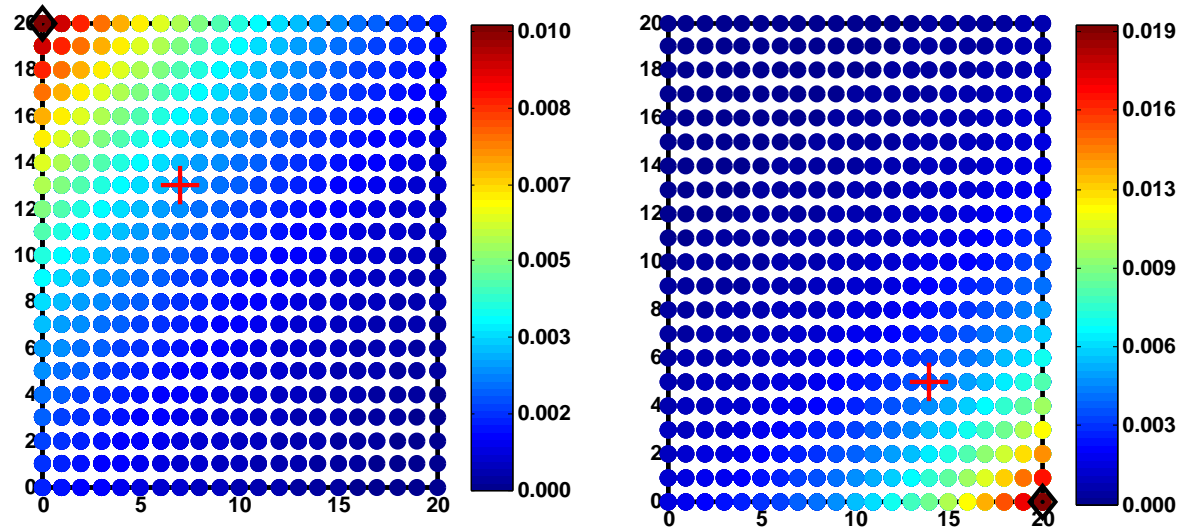
Figure 6: An illustration of degeneracy of exponential models on non-graph data. Colorcoded is the pmf over 2D grid spaces for estimated MLEs *(left plot)* $\boldsymbol{\theta}_{MLE} = (-0.086, 0.086)$ & *(right plot)* $\boldsymbol{\theta}_{MLE} = (0.120, -0.160)$. "+" sign indicates the observed feature and its mean, while the "⋄" shape indicates ERGM mode.

---

[4]We observed the same *ERGM* behavior in separate experiments on other types of features spaces *e.g. Edge-vs-2Star, Edge-vs-GWED, Triangle-vs-GWESP* (GWED: geometrically weighted degree distribution, GWESP: geometrically weighted edgewise shared partners [9].)

Algorithm 1 Outline of `ELSRGM` Procedure

Require: Given an example graph $G^\star \in \mathcal{G}_n$ and a feature vector function $\boldsymbol{f} : \mathcal{G}_n \to \mathbb{R}^d$:

1: Generate a neighborhood $\mathcal{B} \subseteq \mathcal{G}_n$ by performing a random walk in $\mathcal{G}_n$ starting at $G^\star$.
2: Compute the set of statistics $\mathcal{P}(\mathcal{B})$.
3: Compute a mapping $\boldsymbol{e} : \mathbb{R}^d \to \mathbb{S}^{p-1}$ to embed $\mathcal{P}(\mathcal{B})$ onto $\mathbb{S}^{p-1}$, a surface of a sphere in $\mathbb{R}^p$.
4: Estimate the parameters for the von Mises-Fisher density f over the space of hyper-spherical features $\mathbb{S}^{p-1}$.
5: Approximate f with a density $\hat{f}$, a mixture of kernels centered around the spherical features corresponding to graphs in $\mathcal{B}$ and recompute $\mathcal{B}$ when new sample graphs are discovered.

---

3 Exponential Locally Spherical Random Graph Model

The main result of section 2 provides us with a very important insight, that is; type II degeneracy in $ERGMs$ is due to the bounded nature of discrete exponential models. The formulation of such models is sensitive to the geometry of the support space $\mathcal{P}$. The geometry of the convex hull on $\mathcal{P}$ defines which points are likely to have *most* or *all* of the probability mass placed on them. If $\boldsymbol{x} \notin \mathrm{rbd}(C)$ then a model computed for $\boldsymbol{x}$ will place very little probability mass on $\boldsymbol{x}$, while most of the mass is placed on some point *(i.e. mode)* $\boldsymbol{x}^\star \in \mathrm{rbd}(C)$. This result suggests that mapping all points $\boldsymbol{x} \in \mathcal{P}$ onto a surface that belongs to its relative boundary and defining a model in the new space would solve the degeneracy issue. In the following sections, we propose mapping all observed features onto a spherical surface $\mathbb{S}^{p-1}$ since every point on it belongs to $\mathrm{rbd}(C(\mathbb{S}^{p-1}))$. We then define a distribution and sampling techniques for graphs over the resulting feature space.

3.1 Algorithm

In this section we describe the new model for graph sampling, one that can be used to generate non-degenerate graphs similar to the given one. Since the approach we are proposing is based on an exponential family model over the locally spherical embeddings of graph features, we will call it an Exponential Locally Spherical Random Graph Model, or `ELSRGM` for short. Our approach is executed in several steps as outlined in Algorithm 1.

First, we sample the neighborhood $\mathcal{B} \subseteq \mathcal{G}_n$ around the given example graph $G^\star$, and then compute the set $\mathcal{P}(\mathcal{B}) = \{\boldsymbol{f}(G) : G \in \mathcal{B}\}$ of feature vectors corresponding to the graphs in the neighborhood. If the space of graphs ($\mathcal{G}_n$ or its subset) is small, then it may be possible to consider all graphs in the set. Otherwise, the neighborhood

---

Algorithm 2 Outline of Spherical Embedding

Require: Given dissimilarity matrix $D_{n \times n}$, with n number of graphs.

1: If the spherical point positions are given by $X_i,\ i = 1, \cdots, n$, then $\langle X_i, X_j \rangle = r^2 \cos \beta_{ij}$, with $\beta_{ij} = \frac{d_{ij}}{r}$.

2: If X in unknown, compute for X such that $XX^T = Z$, where $Z_{ij} = r^2 \cos \beta_{ij}$ and $d_{ij} \in D$. Find the radius of sphere as $r^\star = \arg\min_r \lambda_1\{Z(r)\}$. Where λ_1 is the smallest eigenvalue of Z.

3: Set $\hat{Z} = \frac{Z}{r^\star}$ and $X^\star = \arg\min_{X, x^T x = 1} \|XX^T = \hat{Z}\|$

4: Decompose $\hat{Z}$, $\hat{Z} = U\Lambda U^T$. Set the embedding positional matrix to be $X^\star = U_{n \times k}\Lambda_{k \times k}^{1/2}$, where k is chosen such that the elements of $U_{n \times k}$ corresponds to the largest k eigenvalues of $\Lambda_{k \times k}$.

---

of $G^\star$ is sampled by a random walk which at each step considers graph one edge deletion/insertion away from the current graph.

The resulting feature set is then embedded in a $p-1$ dimensional unit hypersphere $\mathbb{S}^{p-1} \subset \mathbb{R}^p$ by a linear mapping $\boldsymbol{e} : \mathbb{R}^d \to \mathbb{S}^{p-1} \subset \mathbb{R}^p$. We are using the spherical embedding approach of [18], Algorithm 2), in which the mapping is chosen to minimize the Frobenius matrix distance between the normalized dissimilarity matrix (in our case, matrix of Euclidean distances between feature vectors in $\mathcal{P}(\mathcal{B})$) and the matrix of the Euclidean outer product (in $\mathbb{R}^p$) for the vectors in $\{\boldsymbol{e}(\boldsymbol{x}) : \boldsymbol{x} \in \mathcal{P}(\mathcal{B})\}$. We denote the set of resulting spherical features by $\mathcal{S}(\mathcal{B}) = \{\boldsymbol{e}(\boldsymbol{f}(G)) : G \in \mathcal{B}\}$.[5] One of the beneficial properties of such embedding is that it preserves neighborhood properties, i.e., transformed feature vectors close to each other are mapped to vectors which are also close to each other. For our case, the embedding is *locally* spherical as we determine the mapping based only on a small subset of possible observed graphs, and determining the spherical coordinates for the rest by recomputing the embedding iteratively based on the uncovering of new candidate graphs for the neighborhood $\mathcal{B}$. As a result, the distance preserving property may hold only for the graph neighborhood on which the transformation was estimated.

The next step is to estimate a distribution over all possible values in $\mathbb{S}^{p-1}$ of the spherical features. We propose to use von Mises-Fisher directional distribution (denoted in the rest of the paper as VMFD) with a pdf

$$f_{vmfd}(\boldsymbol{y}|\boldsymbol{\mu}, \kappa) = \frac{\kappa^{\frac{p}{2}-1}}{(2\pi)^{\frac{p}{2}} I_{\frac{p}{2}-1}(\kappa)} \exp\left(\kappa\boldsymbol{\mu}^T\boldsymbol{y}\right),\ \boldsymbol{y} \in \mathbb{S}^{p-1} \tag{7}$$

where $\boldsymbol{\mu}$ is the location parameter, $\kappa \geq 0$ is the concentration parameter, and I_k denotes the modified Bessel function of the first kind and order k. VMFD is a member of the exponential family; unlike a general exponential family distribution, it is symmetric with $\boldsymbol{y} = \boldsymbol{\mu}$ serving as both its mean and mode. Parameter κ determines how concentrated the density is around the mode. When $\kappa = 0$, VMFD corresponds

---

[5] We refer to the spherically embedded graph feature vectors as *coordinates* of the graph.

to the uniform distribution over the hyper-sphere $\mathbb{S}^{p-1}$, while as $\kappa \to \infty$, VMFD is concentrated at the point $\boldsymbol{\mu}$. See [12] for more details on VMFD. As we wish the example graph $G^\star$ to be the mode for the distribution over possible graphs, we set the mode of VMFD to $\hat{\boldsymbol{\mu}} = \boldsymbol{e}\left(\boldsymbol{f}\left(G^\star\right)\right)$. Using MLE, one can compute for $\hat{\kappa}$ from $\mathcal{S}\left(\mathcal{B}\right)$ with already set location parameter $\hat{\boldsymbol{\mu}}$i.e. $\hat{\kappa}_{mle} = \frac{p-1}{2\left(1 - \hat{\boldsymbol{\mu}}^T \sum_{G \in \mathcal{S}(\mathcal{B})} \boldsymbol{e}(\boldsymbol{f}(G))/|\mathcal{S}(\mathcal{B})|\right)}$. Given only a single example graph, the concentration parameter is undefined and also, $\hat{\kappa}_{mle}$ can be observed to be a function of the random-walk-based sampling procedure, which is arbitrary. An alternative is to treat $\hat{\kappa}$ as a user-defined parameter controlling how concentrated the region around the example graph is. As a result, one obtains a distribution with density

$$f\left(\boldsymbol{y}\right) \equiv f_{vmfd}\left(\boldsymbol{y}|\hat{\boldsymbol{\mu}}, \hat{\kappa}\right) \tag{8}$$

over possible spherical feature vectors $\boldsymbol{y} \in \mathbb{S}^{p-1}$.

We note that one can employ a different family of distributions other than VMDF for the task of modeling spherical data. Other candidates provide more degrees of freedom, but also more difficult parameter estimation methods [12]. We leave the investigation of this aspect of our algorithm for future work.

3.2 Sampling under ELSRGM

Using the ideas and steps from the previous subsection, we estimate a probability density function over a unit sphere on the domain of possible spherical feature values for graphs. Before proceeding further, however, we need to resolve several issues. First, we are interested in a discrete probability distribution over a very large but *finite* set of possible features $\mathcal{S}\left(\mathcal{B}\right)$. How would one obtain such a distribution P from the VMFD density? Second, what portion of the mass of f is to be associated with each of the coordinates for possible graphs $G \in \mathcal{G}_n$ (or $\boldsymbol{y} \in \mathcal{S}\left(\mathcal{B}\right)$)? Third, it is infeasible to enumerate all of the possible graphs in $\mathcal{G}_n$, and it may be infeasible to consider all possible spherical feature vectors. How can one perform sampling in $\mathcal{G}_n$ so that the resulting spherical features are distributed according to P? The following subsections will outline our approach to resolving these issues.

3.2.1 Density as an Approximation to the Smoothed Distribution

As a first step, consider a setting where all possible spherical feature vectors $\mathcal{S}\left(\mathcal{B}\right)$ can be enumerated for a fixed number of nodes n. We propose to approximate VMFD using a mixture of kernels with one mixture component for each graph's spherical coordinates. (Alternatively, one can consider one mixture component for each possible spherical feature vector in $\mathcal{S}\left(\mathcal{B}\right)$.) Intuitively, we consider the density f to be a baseline approximation to the distribution obtained by smoothing a discrete distribution over spherical feature values $\mathcal{S}\left(\mathcal{B}\right)$. More formally, for a set of graphs $\mathcal{G}$, we will index its elements $G_1, \ldots, G_K$ with $K = |\mathcal{G}|$, and denote the spherical feature vector for graph G_k by $\boldsymbol{y}_k = \boldsymbol{e}\left(\boldsymbol{f}(G_k)\right) \in \mathcal{S}\left(\mathcal{B}\right)$.

Assume $f(\boldsymbol{y})$ is the estimated density in Equation (8). Let

$$\hat{f}(\boldsymbol{y}|\boldsymbol{\pi}) = \sum_{k=1}^{K} \pi_k h_k(\boldsymbol{y}) \qquad (9)$$

with $h_k(\boldsymbol{y}) \equiv f_{vmfd}(\boldsymbol{y}|\boldsymbol{y}_k, \kappa_{h_k})$, where $\boldsymbol{y}_k$ is the location for kernel h_k, and all of the kernels share a user defined concentration (bandwidth) κ_{h_k} chosen to assign more weight to points closer to $\boldsymbol{y}_k$. The parameters $\boldsymbol{\pi} = \{\pi_k : k = 1, \ldots, K\}$, $\pi_k \geq 0$, $\sum_{k=1}^{K} \pi_k = 1$ are probabilities associated with each $\boldsymbol{y}_k$.

The estimation of the π's is carried out by optimizing with respect to the Kullback-Leibler (KL) information:

$$J(\boldsymbol{\pi}) = KL\left(\hat{f}(\cdot|\boldsymbol{\pi}) \parallel f\right) - \lambda\left(\sum_{k=1}^{K} \pi_k - 1\right) \qquad (10)$$

where λ is the Lagrange multiplier. It can be shown that the objective function $J(\boldsymbol{\pi})$ in (10) is convex, and it can be minimized efficiently using standard convex optimization techniques.

3.2.2 Metropolis-Hastings Algorithm

The goal is to draw samples from $\mathcal{G}_n$ according to the probability mass $\{\pi_1, \ldots, \pi_K\}$, where $K = |\mathcal{G}_n|$. If the number of nodes n is small (≤ 12), random graphs can be enumerated using `nauty` ([13]) and it is possible to identify all possible feature values to compute probability masses $\pi_1, \ldots, \pi_K$ associated with each graph. Graphs can then be sampled directly according to the resulting multinomial distribution. When all possible graphs of n-nodes (for $n \leq 10$) are observed, an alternative is to make use of the Metropolis-Hastings approach as outlined in Algorithm 3. However, in practice, the number of nodes is usually too large to explicitly consider all possible graphs, and the initial neighborhood $\mathcal{B}$ would include only a small portion of all graphs. We propose an approach that will allow us to draw samples from the distribution over graphs including the graphs in $\mathcal{G}_n \setminus \mathcal{B}$.

For better understanding, we first propose a graph generation approach assuming n is small, and all graphs $G_1, \ldots, G_K$ can be enumerated. We will draw a point in the embedding space, and then employ Markov Chain Monte Carlo (MCMC) approach to draw a graph "corresponding" to this point by constructing a Markov chain in the space $\mathcal{G}_n$ of graphs. The pseudocode is presented in Algorithm 3. If we could draw samples directly from $\hat{f}$, this procedure would be equivalent to drawing a vector $\boldsymbol{y}^\star \sim \hat{f}$, and then trying to identify out which of the K components was used to generate $\boldsymbol{y}^\star$ by using MCMC in the posterior over $k = \{1, \ldots, K\}$. The only approximation employed in this case is that instead of drawing $\boldsymbol{y}^\star \sim \hat{f}$, we are drawing $\boldsymbol{y}^\star \sim f$, but according to the KL-divergence measure (Figure 8), $\hat{f}$ and f are very close.

Algorithm 3 MCMC Algorithm for Sampling Graphs

Require: f is given.
1: Draw a sample $\boldsymbol{y}^\star \sim f$. Set $t = 0$, and perturb the example graph $G \in \mathcal{G}_n$ to generate a random graph $G^0 \in \mathcal{G}_n$ to initialize the chain.
2: **repeat**
3:     $t \leftarrow t + 1$.
4:     Sample a proposal graph G' from the proposal distribution at time t, $Q(G'|G^{t-1})$.
5:     Compute ratio, $r = \frac{P(G'|\boldsymbol{y}^\star)Q\left(G^{t-1}|G'\right)}{P(G^{t-1}|\boldsymbol{y}^\star)Q(G'|G^{t-1})}$
     where $P(G = G_k|\boldsymbol{y}) \propto \pi_k h_k\left(\boldsymbol{y}\right)$.
6:     $G^t = \begin{cases} G' & with\ prob\ min(r, 1) \\ G^{t-1} & o.w. \end{cases}$
7: **until** convergence of the chain

To sample from the von Mises-Fisher distribution f we follow the approach outlined in [19]. For proposal distribution $Q\left(\cdot|G^{t-1}\right)$ we consider a uniform distribution over graphs one edge insertion/deletion away from G^{t-1}. In our experiments on small- and large-sized graphs, with the number of iterations set to $T = 1000$, the algorithm is observed to converge and does produce graphs with topologies that resemble the observed graph.

For larger graphs, however, we cannot compute explicitly π_k for all K graphs. This situation can be thought of as similar to the case of *countably infinite* number of objects in which a Dirichlet process can be employed to assign some weight to graphs that are not yet observed. It is as if, f is approximated by a countably infinite mixture $\hat{f}\left(\boldsymbol{y}\right) = \sum_{k=1}^{\infty} \pi_k h_k\left(\boldsymbol{y}\right)$, i.e., assuming that the number of graphs is countably infinite instead of just very large. In this case, we are employing Dirichlet process mixture model $DP\left(\alpha, G_0\right)$, where α is a concentration parameter and G_0 is a uniform distribution $u_p\left(\boldsymbol{\mu}\right) \propto 1$, $\boldsymbol{\mu} \in \mathbb{S}^{p-1}$ for the kernel location since assume that each yet unobserved set of features is equally likely.

Assuming that K different graphs have been observed, equation 9 then becomes

$$\hat{f}\left(\boldsymbol{y}|\boldsymbol{\pi}\right) = \sum_{k=1}^{K} \pi_k h_k\left(\boldsymbol{y}\right) + \frac{\alpha}{K + \alpha} u_p\left(\boldsymbol{y}\right)$$

Algorithm 4 details the pseudocode for sampling of small to large graphs on a spherical space. For basic information on Dirichlet process mixtures and their inference see [14].

4   Experimental Evaluation

We adopt the common goodness of fit measures in network generation studies to investigate how well our model fits the observed data (i.e. by comparing the observed

---

Algorithm 4 MCMC-DP Algorithm for Sampling Graphs

Require: f is given. Graphs $G_1, \ldots, G_L$ initially observed. Set $K = L$. Estimate $\boldsymbol{\pi}$
   by minimizing (10).
1: Draw a sample $\boldsymbol{y}^\star \sim f$. Set $t = 0$, and perturb[6] the example graph $G \in \mathcal{G}_n$ to
   generate a random graph $G^0 \in \mathcal{G}_n$ to initialize the chain.
2: **repeat**
3:    $t \leftarrow t + 1$.
4:    Sample a proposal graph G' from the proposal distribution at time t,
      $Q(G'|G^{t-1})$.
5:    Compute ratio, $r = \frac{P(G'|\boldsymbol{y}^\star)Q(G^{t-1}|G')}{P(G^{t-1}|\boldsymbol{y}^\star)Q(G'|G^{t-1})}$
6:    where
$$P(G = G_k|\boldsymbol{y}) \propto \begin{cases} \pi_k h_k\left(\boldsymbol{y}\right) & k = 1, \ldots, K, \\ \frac{\alpha}{K+\alpha} u_p\left(\boldsymbol{y}\right) & G = new. \end{cases}$$
7:    Set $G^t = \begin{cases} G' & with\ prob\ min(r, 1) \\ G^{t-1} & o.w. \end{cases}$
8:    If G^t is new, then set $K = K + 1$, $G_K = G^t$.

   - Compute $\boldsymbol{f}(G_K)$, set $\mathcal{B} = \mathcal{B} \cup \boldsymbol{f}(G_K)$.

   - Re-compute $\mathcal{S}\left(\mathcal{B}\right)$ (to generalize to new samples) and re-estimate
     $\pi_1, \ldots, \pi_K$.

9: **until** convergence of the chain

---

statistics with a range of the same statistics obtained from simulating many networks
using the fitted model) [9]. If the observed network is not typical of the simulated
networks for a particular measure then the model is either degenerate or simply a
misfit. We first consider the *degree distribution*, which is defined as the statistics:
$D_0, D_1, \cdots, D_{n-1}$, with each D_i representing the number of nodes with i edges con-
nected to them, divided by n. Secondly, we compute the *edgewise shared partner
distribution*, which is defined as the statistics: $EP_0, EP_1, \cdots, EP_{n-2}$, with each EP_i
representing the number of edges in the graph between two nodes that share exactly
i neighbors in common, divided by the total number of edges. Thirdly, we compute
the *triad census distribution*, which is the proportion of $3 - node$ sets having $0, 1, 2,$
or 3 edges among them. Lastly, we compute the *minimum geodesic distance distribu-
tion*; which is the proportion of pairs of nodes whose shortest connecting path is of
length k, for $k = 1, 2, \cdots$. Also, pairs of nodes that are not connected are classified
as $k = \infty$.

   In all our experiments, we define each feature vector to be,

$$\boldsymbol{f}\left(G\right) = \left(f_{edge}, f_{2\star}, \cdots, f_{(11)\star}, f_\triangle\right)^T$$

These features have been observed to be among the set of subgraph patterns that
capture social interactions and network formation in real processes [4].

4.1 Small graphs

We first consider the experiments for $\mathcal{G}_n$ with $n = 8$ nodes. Figure 7 (left) shows a two component 8-node synthetic graph, G_{test_1}, while Figure 8 displays the KL-divergence $KL\left(\hat{f} \parallel f\right)$ computed by first observing 300 8-node graphs (graph-edit distance ≤ 3 of neighborhood around G_{test_1}), and allowing the `MCMC-DP` (Algorithm 4) to discover and fill-in the neighborhood set $\mathcal{B}$, with the baseline $VMFD$ centered at the coordinates of G_{test_1}. G_{test_1} was chosen to test our hypothesis for the extended feature space, that is if $\boldsymbol{x}^\star \in \text{rint}(\bar{C})$ e.g. $\boldsymbol{x}^\star = \boldsymbol{f}\left(G_{test1}\right) \in \text{rint}\left(\bar{\mathcal{C}}\right)$, then the corresponding $ERGM$ will be degenerate in relation to the analysis of Theorem 1. G_{test_2} is used to test our second hypothesis, i.e. if $\boldsymbol{x}^\star \in \text{rbd}(\bar{C})$ then the ERGM specified by $\boldsymbol{x}^\star$ is most likely to generate realistic graphs. In Figure 7 (right) we show an 8-node synthetic graph G_{test_2} whose extended feature vector $(\boldsymbol{x}^\star = \boldsymbol{f}\left(G_{test2}\right) \in \text{rbd}\left(\bar{\mathcal{C}}\right))$.

For the experimental set-up, the bandwidth for each kernel h_k is treated as a user defined parameter and is set to $\kappa_{h_k} = 400$, while the DP prior parameter is set to $\alpha = 0.5$. Parameters for the $VMFD$ are set as discussed in section 3, with κ fixed at 140 to control the concentration of f_{VMFD}. We compute the parameters for the ERGM model P and simulations using the `statnet` package [6].

Figure 9, depicts 100 simulated summary results for $\boldsymbol{f}(G_{test_1}) \in \text{rint}\left(\bar{\mathcal{C}}\right)$. The result obtained from our `ELSRGM` (in unshaded blue box plots), displays no sign of degeneracy while the $ERGM$ result shows summary statistics that appear to relate to those of empty and complete graphs- a sign of degeneracy. Figure 10, shows summary statistics of simulated networks obtained for the $ERGM$ and the `ELSRGM` models when $\boldsymbol{f}(G_{test_2}) \in \text{rbd}\left(\bar{\mathcal{C}}\right)$. The results show no sign of degeneracy from both the specified $ERGM$ and the proposed `ELSRGM` . This confirms our second hypothesis, that is having extended feature vectors lie on the relative boundary of the convex hull enables the generation of realistic graphs from exponential family models.

4.2 Larger Graphs

For larger data sets, we first consider a undirected `Dolphin` social network with 62 nodes [3, 11]. We apply the `MCMCDP` approach as outlined in Algorithm 4, by sampling graphs via a random walk for 1000 steps starting from a random network $G^0 \in \mathcal{B}$. Figure 11 summarizes the results of 100 simulations for the `Dolphin` network from the two approaches. The proposed `ELSRGM` shows a relatively better performance of capturing the distribution of statistics of the `Dolphin` network. It is appears that ERGM model specified by simple statistics is incapable of generating the distribution of statistics that resembles those of the observed network. The lack of fit by an ERGM model in the degree distribution, geodesic distance, and shared partners distribution indicates the presence of degeneracy.

We next evaluate `ELSRGM` on 205-node `Faux-Mesa-High` social network [6, 15]. We again apply the `MCMC-DP` algorithm by sampling graphs via a random walk for 1000 steps starting from a random network $G^0 \in \mathcal{B}$. Figure 12 depicts the results of
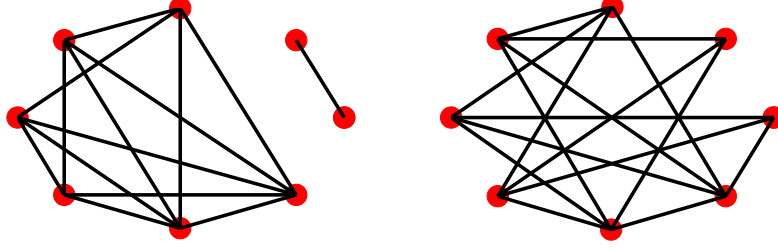
Figure 7: Two 8-node synthetic test graphs. **Left:** 2-component graph G_{test1} falling inside the relative interior of the convex hull $\bar{\mathcal{C}}$ of extended features. **Right:** G_{test2} falling on the relative boundary of the convex hull $\bar{\mathcal{C}}$ for the extended feature space.
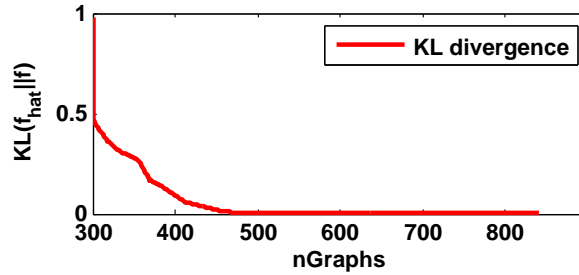

Figure 8: Sample KL-divergence $KL\left(\hat{f} \parallel f\right)$ for `ELSRGM` as more graphs get uncovered. The initial neighborhood $\mathcal{B} \subset \mathcal{G}_n(n = 8)$ is built around G_{test1}.

100 simulated networks from the specified `ELSRGM` and $ERGM$ models. Again, we observe that `ELSRGM` generates distributions of statistics that resemble those of the observed network while ERGMs shows signs of misspecification.

Finally, we assess `ELSRGM`'s performance on a benchmark data set from social network analysis, which is used in testing whether a model can overcome the degeneracy phenomenon. We consider the first matrix (39 nodes) of the Kapferer's tailor shop data [10]. The result of 100 simulated network summaries shown in Figure 13 suggests that the proposed `ELSRGM` does not suffer from degeneracy. ERGM results were not included for this data set due to severe unstable estimation of MLE using simple specifications considered in our analysis (corresponding to the first type of degeneracy).

5 Conclusion and Future Work

In this paper, we investigated the cause of degeneracy in ERGMs, and explained the degeneracy as related to the feature vectors belonging to the relative interior of the convex polytope of feature values. To correct this issue, we proposed an algorithm that
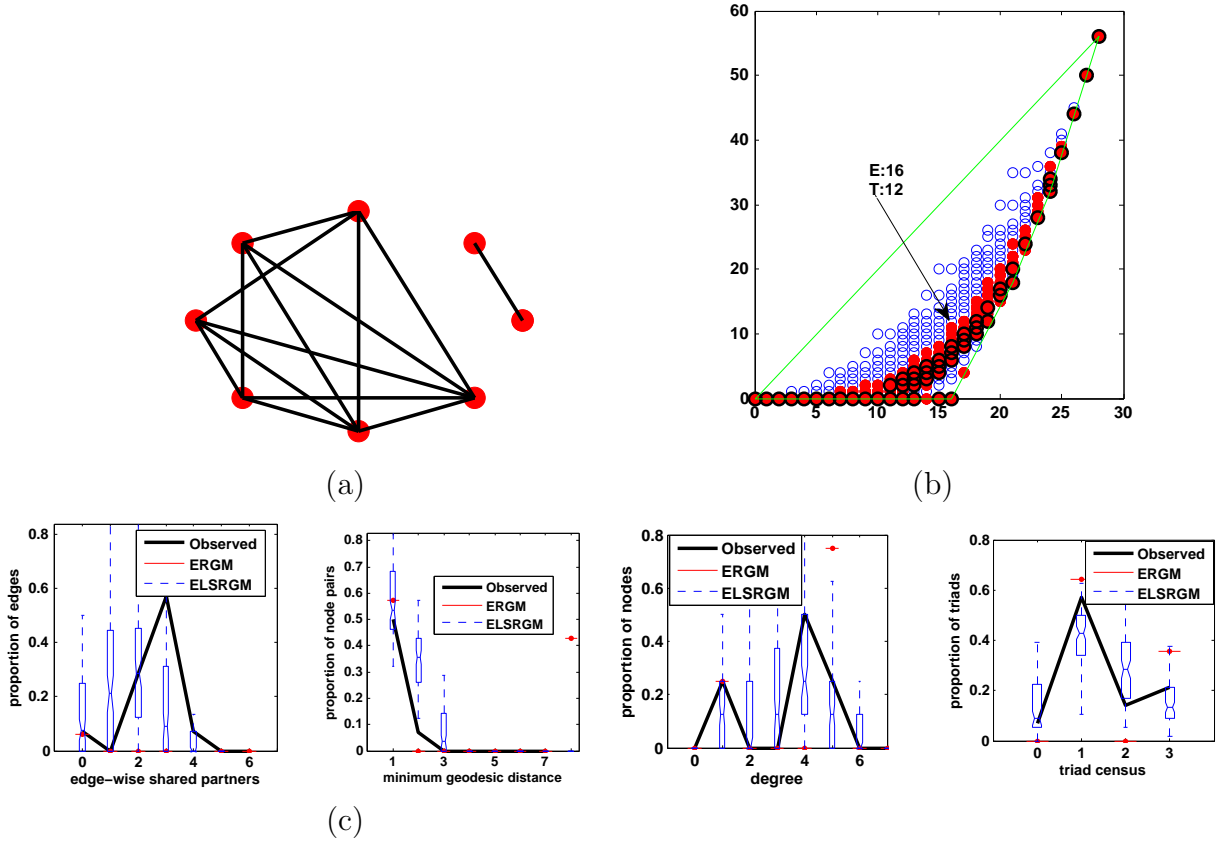
Figure 9: (a) Synthetic graph G_{test_1}; (b) Corresponding feature-pair: $\boldsymbol{f}(G_{test_1}) \in$ rint$(\bar{\mathcal{C}})$; (c) Simulations from models given G_{test_1}. The observed statistics are indicated by the solid lines; the box plots include the median and interquartile range of simulated networks. *ERGMs* show low variance and all 100 samples seem to be placed on the same network- sign of degeneracy.

uses spherical features as points on a sphere would belong to the relative boundary of the corresponding convex set rather than its relative interior. Based on this mapping, we outlined a novel model for graphs (`ELSRGM`) and an approach to sampling graphs from this model. In several synthetic and real-world social network, our approach generated graphs with statistics similar to that of the example graphs while not suffering from the issue of degeneracy (unlike ERGMs).

`ELSRGM` opens up a new class of graph sampling models: those based on spherical features. In this paper, we made several modeling choices, e.g., von Mises-Fisher distribution for spherical density, kernel approach assigning mass to individual feature vectors; other modeling choices could also lead to models preserving properties of the example graphs, and they need to be investigated. The insight from the geometric interpretation of the feature vectors realizable as modes may also lead to other types of features (non-spherical) which can lead to models generating realistic graphs.
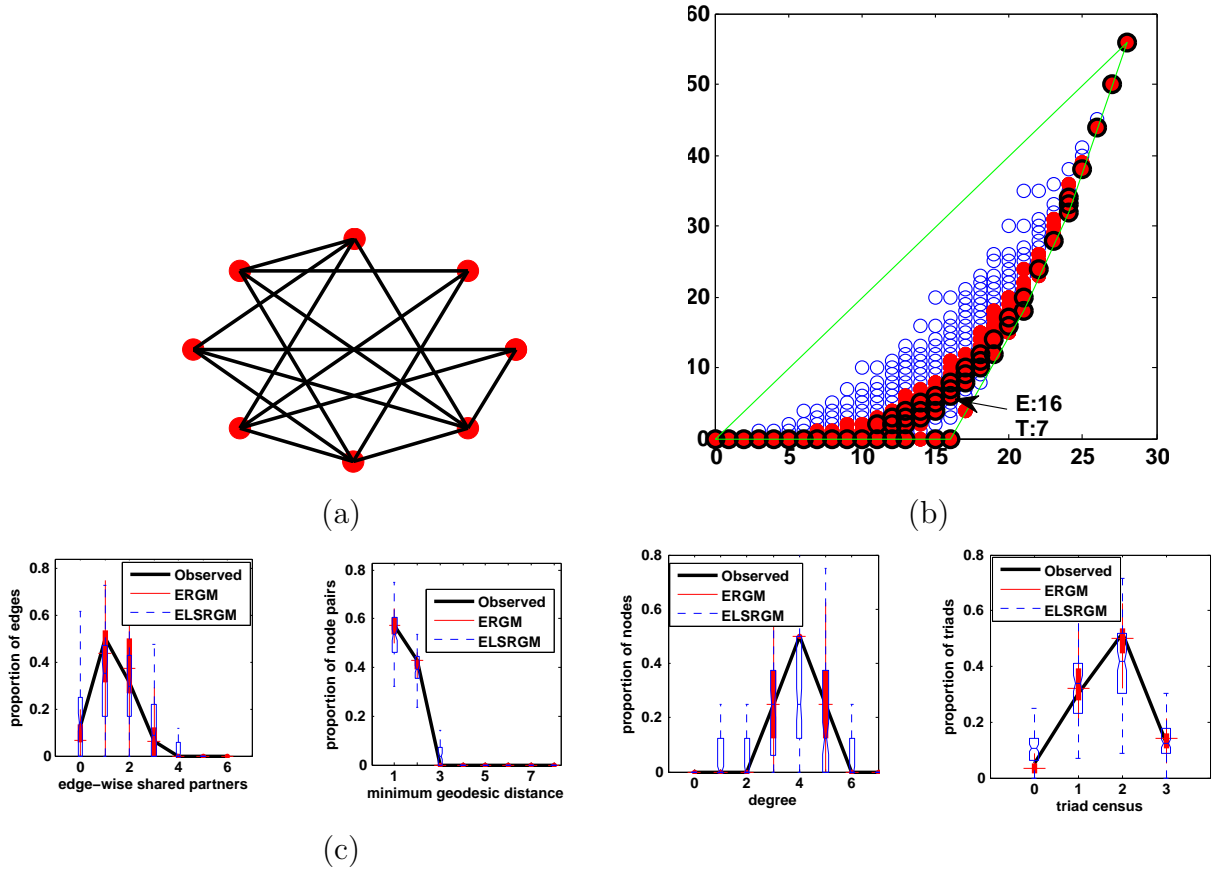
Figure 10: (a) Synthetic graph G_{test_2}; (b) Corresponding feature-pair: $\boldsymbol{f}(G_{test_2}) \in$ rbd $(\bar{\mathcal{C}})$; (c) Simulations from models given G_{test_2}. The observed statistics are indicated by the solid lines; the box plots include the median and interquartile range of simulated networks. Both models are non-degenerate.

6 Acknowledgements

References

[1] O. Barndorff-Nielsen. *Information and exponential families in statistical theory.* Wiley, New York, 1978.

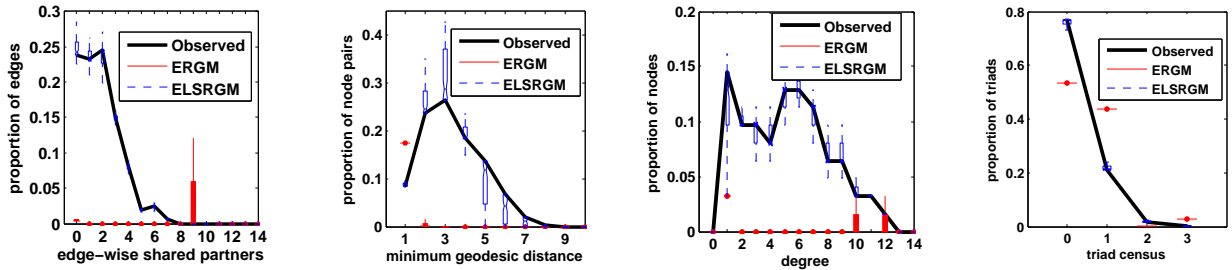[2] A. Brøndsted. *An Introduction to Convex Polytopes.* Springer-Verlag, 1983.

Figure 11: `Dolphin` 62-node network summaries. The observed statistics are indicated by the solid lines; the box plots include the median and interquartile range of simulated networks. *ERGM* plots shows signs of degeneracy effect.
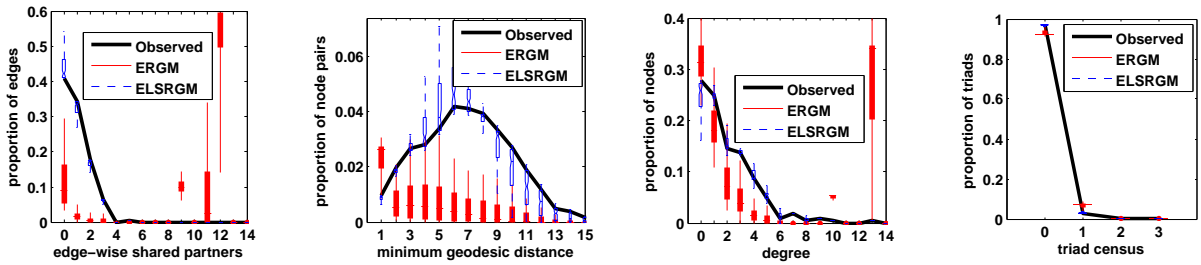

Figure 12: `Faux-Mesa-High` 205-node network summaries. The observed statistics are indicated by the solid lines; box plots summarize the statistics for the simulated networks. *ERGM* plots shows signs of degeneracy effect.
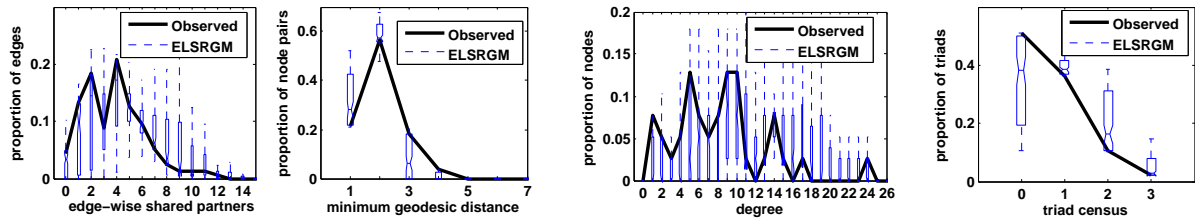

Figure 13: Kapferer's 39-node network summaries. The observed statistics are indicated by the solid lines; box plots summarize the statistics for the simulated networks using `ELSRGM`.

[3] C. L. DuBois and P. Smyth. UCI network data repository, 2008. http://networkdata.ics.uci.edu.

[4] O. Frank and D. Strauss. Markov graphs. *JASA*, 81(395), September 1986.

[5] M. S. Handcock. Assessing degeneracy in statistical models of social networks. Technical Report 39, Center for Statistics and the Social Sciences, University of Washington, 2003.

[6] M. S. Handcock, D. R. Hunter, C. T. Butts, S. M. Goodreau, and M. Morris. `statnet`: Software tools for the representation, visualization, analysis and simulation of network data. *Journal of Statistical Software*, 24(3), 2008.

[7] P. W. Holland and S. Leinhardt. An exponential family of probability distributions for directed graphs (with discussion). *Journal of American Statistical Association*, 76(373), 1981.

[8] D. R. Hunter and M. Handcock. Inference in curved exponential family models for networks. *Journal of Computational and Graphical Statistics*, 15(2), 2006.

[9] D. R. Hunter, M. S. Handcock, C. T. Butts, S. M. Goodreau, and M. Morris. `ergm`: A package to fit, simulate and diagnose exponential-family models for networks. *Journal of Statistical Software*, 24(3), 2008.

[10] B. Kapferer. *Strategy and transaction in an African factory.* Manchester: Manchester University Press., 1972. available from http://www.casos.cs.cmu.edu/computational_tools /datasets/external/kaptail/index2.html.

[11] D. Lusseau, K. Schneider, O. J. Boisseau, P. Haase, E. Slooten, and S. M. Dawson. The bottlenose dolphin community of Doubtful Sound features a large proportion of long-lasting associations. *Behavioral Ecology and Sociobiology*, 54, 2003.

[12] K. V. Mardia and P. E. Jupp. *Directional Statistics.* Wiley, 2000.

[13] B. D. Mckay. Practical graph isomorphism. *Congressus Numerantium*, 30, 1981.

[14] R. M. Neal. Markov chain sampling methods for Dirichlet process mixture models. *Journal of Computational and Graphical Staitistics*, 9(2):249–265, June 2000.

[15] M. D. Resnick and coauthors. Protecting adolescents from harm. findings from the national longitudinal study on adolescent health. *Journal of American Medical Association*, 278(8), 1997.

[16] A. Rinaldo, S. E. Fienberg, and Y. Zhou. On the geometry of discrete exponential families with application to exponential random graph models. *Electronic Journal of Statistics*, 3:446–484, 2009.

[17] S. Wasserman and P. Pattison. Logit models and logistic regression for social networks: An introduction to Markov graphs and $p^\star$ model. *Psychometrii*, 61(3), September 1996.

[18] R. C. Wilson, E. R. Hancock, E. Pekalska, and R. P. W. Duin. Spherical embeddings for non-Euclidean dissimilarities. In *CVPR-10*, pages 1903–1910, June 2010.

[19] A. T. Wood. Simulation of the von Mises-Fisher distribution. *Communications in Statistics - Simulation and Computation*, 23(1), 1994.
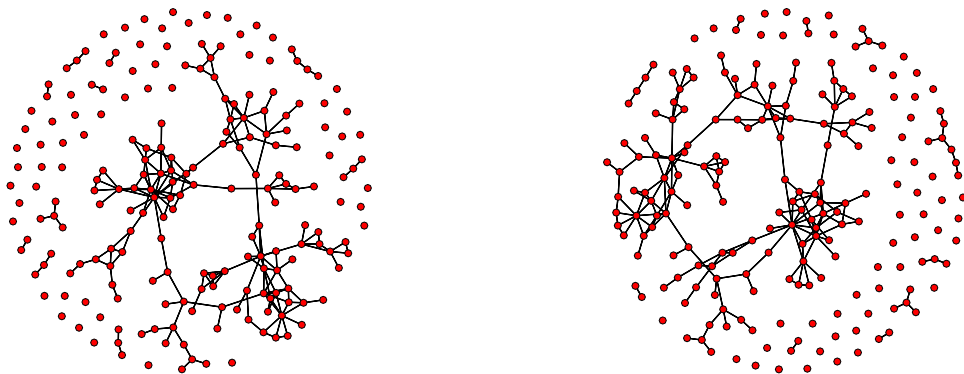
Figure 14: *left:* `Faux-Mesa-High` 205-node original network. *right:* `ELSRGM` generated network.