Sparse Exponential Random Graphs

Suman Chakraborty, Remco van der Hofstad, Frank den Hollander

13 October 2020

Exponential Random Graph Models

ERGM are widely used to capture typical behaviour observed in complex networks. For example, many observed networks are sparse but contain atypical numbers of triangles. More precisely, in a network with n nodes one often observes both order n many edges and order n many triangles. This renders a vast class of random graph models useless for modelling.

To overcome this problem, ERGM proposes to "tilt the random graph exponentially" in order to favour a desired outcome. For example, one might consider the following probability distribution on the space of all graphs on n vertices:

$$P(G) \propto \exp\left(\beta_1 E(G) + \frac{\beta_2}{n} T(G)\right) \tag{1}$$

with $E(G)$ and $T(G)$ the number of edges and triangles in the graph G, respectively. The interpretation is that for $\beta_1 < 0$ and $\beta_2 > 0$ this distribution 'favours' configurations with less edges and more triangles. Unfortunately, it is now known whether in the dense setting (i.e., order n^2 many edges) the above model serves the desired purpose. More precisely, there are different values of the parameter (β_1, β_2) such that typical observations from this model are indistinguishable when n is sufficiently large (in fact, sometimes typical observations are indistinguishable from i.i.d. random graphs). This problem occurs for general classes of models in the dense settings [1, 2, 3].

Although most applications are concerned with the sparse setting [6], few rigorous results are known also in this case. Recently, Mukherjee [5] obtained results in the sparse setting when the sufficient statistics (the expression in the exponent in (1)) is a function of the degrees of the vertices.

We propose to discuss the following problems:

1. What happens if we consider triangles or functions of triangles? (The number of triangles is not a function of the degrees.) For example, $T_f(G_n) = \sum_{i=1}^{n} f(T(G_n, i))$, where $T(G_n, i)$ is the number of triangles in the graph G_n that contain the vertex i. Can we determine a class of functions f such that we can determine the normalising constant for this model?

2. More general statistics have been proposed in the literature [4], such as alternating k-triangles [7] (see below).

3. In which situations does the model produce trivial results? For example, for which statistics are typical observations of the model either complete graphs or empty graphs?

4. When is the model is statistically sound, i.e., are the parameters efficiently identifiable?

5. Can we talk about the limit of these graphs as $n \to \infty$ in the sparse setting (in some suitable sense)? What is the appropriate choice for β_1, β_2 to make the model sparse? (This is also related to the previous items.) Initial computations suggest that β_2 might need to grow like $\log n$ for the model to become sparse.

Alternating k-triangle

A k-triangle with base (a, b) is a graph on $k + 2$ vertices with vertex set $\{a, b, v_1, \ldots, v_k\}$ and edge set

$$(a, b), \qquad (a, v_1), \ldots, (a, v_k), \qquad (b, v_1), \ldots, (b, v_k). \tag{2}$$

Note that for $k \geq 2$ the number of k-triangles in a graph can be written as

$$T_k = \sum_{i<j} a_{ij} \binom{L_{2ij}}{k}, \tag{3}$$

where L_{2ij} is the number of paths between i and j (i.e., the number of common neighbours of i and j), and $a_{ij} = 1$ when there is an edge between i and j, and $a_{ij} = 0$ otherwise. For $k = 1$, on the other hand,

$$T_1 = \frac{1}{3} \sum_{i<j} a_{ij} L_{2ij}, \tag{4}$$

where the factor $\frac{1}{3}$ is to adjust for the 3 times over counting in the configuration. The alternating k-triangle statistics is defined as

$$T(G) = 3T_1 - \frac{T_2}{\lambda} + \frac{T_3}{\lambda^2} + \cdots + (-1)^{(n-2)} \frac{T_{n-3}}{\lambda^{n-3}}, \tag{5}$$

where $\lambda > 0$ is a parameter.

References

[1] S. Bhamidi, G. Bresler, and A. Sly, *Mixing time of exponential random graphs*, The Annals of Applied Probability, 21 (2011), pp. 2146–2170.

[2] S. Bhamidi, S. Chakraborty, S. Cranmer, and B. Desmarais, *Weighted exponential random graph models: Scope and large network limits*, Journal of Statistical Physics, 173 (2018), pp. 704–735.

[3] S. Chatterjee and P. Diaconis, *Estimating and understanding exponential random graph models*, The Annals of Statistics, 41 (2013), pp. 2428–2461.

[4] M. S. Handcock, G. Robins, T. Snijders, J. Moody, and J. Besag, *Assessing degeneracy in statistical models of social networks*, tech. rep., Citeseer, 2003.

[5] S. Mukherjee, *Degeneracy in sparse ergms with functions of degrees as sufficient statistics*, Bernoulli, 26 (2020), pp. 1016–1043.

[6] G. Robins, P. Pattison, Y. Kalish, and D. Lusher, *An introduction to exponential random graph (p*) models for social networks*, Social networks, 29 (2007), pp. 173–191.

[7] T. A. Snijders, P. E. Pattison, G. L. Robins, and M. S. Handcock, *New specifications for exponential random graph models*, Sociological methodology, 36 (2006), pp. 99–153.