1.2 Phase Transitions in Random Graphs

Tejas Iyer, Lukas Lüchtrath, and Elena Magnanini

Many complex systems, in areas as diverse as biology, sociology, and computer science, can be represented as *networks*. That is, the system can be modeled by *nodes* and interactions between them, represented by *edges*. Such systems include, for example, the internet, where nodes are webpages that are connected by hyperlinks (Figure 1), networks in molecular biology (Figure 2), social networks, and communication networks. More generally, large sets of correlated data are represented by networks.

Fig. 1: *Visualization of the internet by Barrett Lyon, Opte project (2003)*

In order to gain a better understanding of certain network effects, for instance the fast spread of news through social media, it is of great importance to identify and study typical structures and features of networks arising in the real world. One easily observable property is the *degree distribution*. Here, the *degree* of a node is the number of other nodes that it is connected to, and the degree distribution then gives the probability of observing a certain degree when picking a node randomly from the network. Interestingly, it has been observed that in many real-world networks, the probability of choosing a node of degree k behaves roughly like $k^{-\tau}$ for some $\tau > 2$, a property known as being *scale free*. This property indicates the existence of exceptionally high degree nodes, thus, informally, indicating that there is no characteristic "scale" in the degrees (Figure 3). Another structural property of many networks is that they tend to display a high degree of *clustering*. Clustering refers to the feature that two nodes connected to a common node are more likely to be connected to each other, and is often measured in terms of the number of *triangles* in the network. Another property of interest is the existence of a connected component that contains a fraction of all the nodes, a *macroscopic connected component*.

Fig. 2: *Visualization of a biological network by Fozail Ahmad, Bioinformatics Review (2016)*

Appropriate mathematical models, for which the emergence of the properties described above are proven rigorously, can be used as null models or benchmarks when it comes to testing algorithms and statistical methods for real-world applications. Moreover, these mathematical proofs can often provide insights into the reasons underlying the emergence of such properties.

Phase transitions in random graphs. In order to incorporate the uncertainties arising in real-world applications, networks are generally modeled mathematically as *random graphs*. Here, the edges and sometimes also the nodes are *random*. Of particular interest is whether or not changing certain parameters leads to dramatic changes in the graph structure, a phenomenon usually referred to as a *phase transition*. This is important since parameters associated with these networks may fluctuate (due to, for example, phenomena such as epidemics arising on these networks), and we want to know under what circumstances these fluctuations will have dramatic consequences.

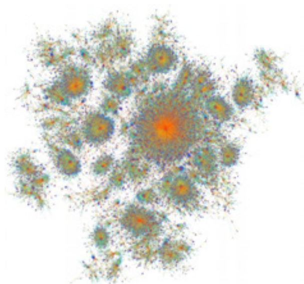

Fig. 3: *Visualization of a scale-free network by Pim van der Hoorn, Networkpages (2020)*

Our research. Our research in Research Group RG 5 *Interacting Random Systems* and the Leibniz Group DYCOMNET *Probabilistic Methods for Dynamic Communication Networks* investigates the emergence and nature of phase transitions in multiple contexts. The first part reports on the emergence of *condensation* in inhomogeneous *preferential attachment models* (popular time-evolving models producing scale-free random graphs), where a positive proportion of the edges in the network may accumulate around nodes of large degree. The second part highlights work related to the presence or absence of macroscopic connected components in *spatial* models, a property that is particularly interesting in the context of wireless telecommunication. The third part is related to work regarding phase transitions in the edge density of exponential random graphs. The latter arise as models incorporating clustering in diverse contexts.

Inhomogeneous preferential attachment models

A popular class of models that displays some of the features associated with complex networks, in particular the property of being *scale free*, is known as *preferential attachment*. Informally, these are sequences of graphs evolving in discrete time, where nodes arrive at discrete time-steps and connect to existing nodes with probability proportional to their degree. Models of this type date back to Yule, 1924, but were popularized in the context of random networks by Barabási and Albert, 1999.

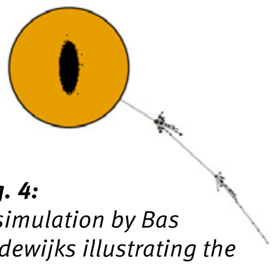

Fig. 4: *A simulation by Bas Lodewijks illustrating the extreme-condensation phase, when $f(i,w) = i^p + w$, $p > 1$. The parameter regime simulated corresponds to that when there is a single node of infinite degree, which may be the large orange node.*

Despite their success, a shortcoming of the classical preferential attachment models is that they fail to encapsulate the inherent *inhomogeneities* arising in real-world networks. For example, in the classical models the oldest nodes will tend to have the largest degrees, whilst on the other hand, in contexts such as the internet, one may expect newer nodes to compete with older ones. Extensions and newer variants have addressed this issue by assigning positive weights to nodes, so that newer nodes attach to previous ones according to a function of their degree and their weight. If this function is monotone increasing in the weight variable, one may regard the weight as the attractiveness of a node. When this function is given by the product of the degree and weight, this model is known as *preferential attachment with multiplicative fitness*, or the *Bianconi–Barabási* model (Bianconi and Barabási, 2001). Here, researchers observed that, when nodes are assigned *independent, identically distributed* (i.i.d.) weights, there is a critical condition on the weight distribution leading to a *condensation* phase transition. In this context, condensation means that a positive fraction of edges in the network accumulates around nodes of maximum weight. This observation was first proved mathematically in Borgs et al., 2007.

Often, due to simplicity, one considers evolving tree models since one expects many properties, such as the degree distribution, to be similar to models involving evolving graphs. A natural framework of evolving trees, which encompasses and generalizes many of the existing models above, posits that nodes v arrive one at a time and are assigned a random i.i.d. *weight* W_v. These weights may take values in an arbitrary measure space $(S, \mathcal{S})$. Newly arriving nodes then connect to a single existing node with probability proportional to a general, measurable fitness function $f : \mathbb{N}_0 \times S \to [0, \infty)$ that incorporates information about the current degree of the target node, and its weight. This model class possesses a rather rich structure. The condensation behavior may roughly be classified according to the following conjectured phases [3]:

1. *Non-condensation phase:* There exists $\lambda > 0$ such that $\sum_{j=1}^{\infty} \mathbb{E}\left[\prod_{i=0}^{j-1} \frac{f(i,W)}{f(i,W)+\lambda}\right] = 1$.
2. *Condensation phase:* For any $\lambda > 0$ such that the sum converges $\sum_{j=1}^{\infty} \mathbb{E}\left[\prod_{i=0}^{j-1} \frac{f(i,W)}{f(i,W)+\lambda}\right] < 1$.
3. *Extreme-condensation phase:* For any $\lambda > 0$, $\sum_{j=1}^{\infty} \mathbb{E}\left[\prod_{i=0}^{j-1} \frac{f(i,W)}{f(i,W)+\lambda}\right] = \infty$.

This conjecture is proved in a number of specific cases in [3]. Both the second and third phase are of particular interest in exploring more detailed properties of the process. In the third phase, all of the mass of edges concentrates on a *sub-linear* number of nodes of large weight and degree. Ongoing work in exploring this phase has led to some more interesting results concerning the limiting infinite tree associated with the model, roughly establishing two other phases:

1. The *no-sideways explosion* phase: For *almost all* W, $\sum_{i=0}^{\infty} \frac{1}{f(i,W)} = \infty$. In this case, under another critical condition, there is either a single infinite path in the associated infinite tree, with every node having finite degree, or, when $f(i,W) > 0$ for all $i \in \mathbb{N}$, there exist infinitely many nodes of infinite degree and uncountably many infinite paths. The former phase may be regarded as an extreme effect of competition on the structure of the infinite tree, making everyone poor in the sense that their degree is finite rather than infinite. In [5], we derived sufficient criteria for either case.
2. The *certain-sideways explosion* phase: For almost all W, $\sum_{i=0}^{\infty} \frac{1}{f(i,W)} < \infty$. In this case, there are precisely two scenarios: Either the infinite tree contains exactly one node of infinite degree or exactly one infinite path (but not both!); see Figures 4 and 5. We also derived sufficient conditions for either scenario, proving a phase transition in many particular instances of the model [4].

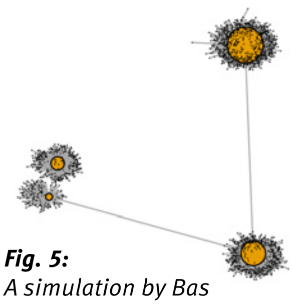

Fig. 5: *A simulation by Bas Lodewijks illustrating the "extreme condensation" phase, when $f(i,w) = i^p + w$, $p > 1$. In the parameter regime shown, there is a single infinite path.*

Work on this problem is closely related to so-called *Crump–Mode–Jagers branching processes*, which model the size and structures of populations in continuous times (for example, the number of species of a biological entity) and are thus of interest in regards to other potential applications outside network science.

Percolation phase transition in the weight-dependent random connection model

The models described in the previous part can produce scale-free networks. By further embedding the nodes into *space*, in addition to assigning them weights, we can also incorporate clustering. More precisely, in the models we consider in this part, pairs of nodes that are located *closer* to each other are more likely to connect, in addition to pairs where one node has a high weight. A well-known example is the *Boolean model*, where the weights are random radii, and two vertices are connected when the associated balls centered at the nodes intersect (Figure 6). Unlike the model of the previous part, this model has no time parameter and consists of infinitely many nodes homogeneously distributed in the entire Euclidean space. A main interest of DYCOMNET lies in finding criteria for the presence or absence of infinite connected components. For communication applications, the components are the parts of the network through which messages can be exchanged. Hence, the existence of infinitely large components is of fundamental importance. We say that a graph with an infinite component *percolates*. Originally, *percolation* was introduced by Broadbent and Hammersley (1957). The idea was to model a porous medium as a random graph, and an infinite component is interpreted as a fluid being able to percolate through the medium.

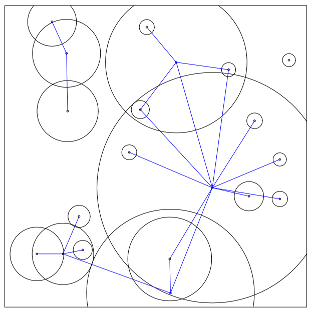

Fig. 6: *Realization of the Boolean model based on a Poisson point cloud*

A key quantity in our setting is the amount of long edges as a measure of what different regions a vertex can reach. Formally, the nodes are embedded into d-dimensional Euclidean space through a standard *Poisson point process* and additionally assigned an independent *mark* distributed uniformly on $(0,1)$ representing its *inverse weight*. Any pair of nodes x and y with marks u_x

and u_y then independently forms an edge with probability $\rho(\beta^{-1} g(u_x, u_y)|x - y|^d)$. Here, ρ is a non-increasing function and, hence, short distances lead to larger probabilities. The function g is assumed to be non-increasing such that smaller marks (i.e., larger weights) lead to higher connection probabilities. Additionally, $\beta > 0$ is an intensity parameter scaling the expected degree. Many established models belong to the presented class. For instance, the aforementioned Boolean model is given by the choice $\rho = \mathbf{1}_{[0,1]}$ and $g(u_x, u_y) = (u_x^{-\gamma/d} + u_y^{-\gamma/d})^{-d}$ for some $\gamma \in (0, 1)$.

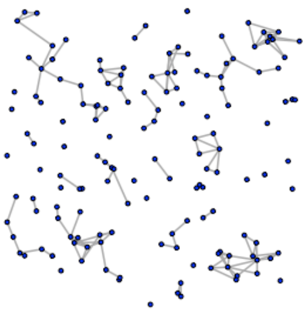

Fig. 7: Snapshot of a weight-dependent random connection model with $\delta_{\text{eff}} > 2$ and small β

As it is well known that for $d \geq 2$ graphs of this class contain an infinite connected component for large values of β, we are interested in whether there is a phase transition such that an infinite component no longer exists when β is small. The key to answering this question is to quantify the occurrence of long edges on various scales via the *effective decay exponent* defined as

$$\delta_{\text{eff}} := - \lim_{n \to \infty} \frac{\log \int_{1/n}^{1} \int_{1/n}^{1} \rho(\beta^{-1} g(s,t) n) \, \mathrm{d}s \, \mathrm{d}t}{\log n}.$$

It turns out that $\delta_{\text{eff}} > 2$ is a sufficient condition for the absence of percolation for small enough β (Figure 7). Additionally, one can express in terms of δ_{eff} the decay of the probability that the cardinality and/or the spatial extension of a typical component in the non-percolation regime exceed a certain size [6]. The situation in dimension $d = 1$ is different. Here, the existence of an infinite component is rather hard to achieve. Nevertheless, it turns out that $\delta_{\text{eff}} < 2$ implies the presence for large β and conversely $\delta_{\text{eff}} > 2$ implies the absence of an infinite component for all β. We also show the existence of a largest component of linear size in larger and larger snapshots of the graph [2]. The idea behind the results is that $n^{-\delta_{\text{eff}}}$ approximates the probability of an edge between two vertices chosen uniformly among two sets of n vertices at distance roughly $n^{1/d}$ when n is large. Ignoring correlations, there are n^2 trials to form an edge between those sets and, therefore, the probability of connecting two distant sets of nodes increases for $\delta_{\text{eff}} < 2$, but decreases for $\delta_{\text{eff}} > 2$.

Limit theorems for exponential random graphs

Fig. 8: Snapshot of a weight-dependent random connection model with $\delta_{\text{eff}} < 2$ and small β

Exponential random graphs are another ubiquitous class of models that can incorporate clustering amongst other network tendencies. From the point of view of sociology, one of the main desires is to understand how the connectivity in local communities can influence the overall network structure. This can be modeled by considering a probability distribution that biases the occurrence of certain features, such as the number of edges or *triangles*, and then analyzing the large-scale properties of random networks sampled according to this distribution. Mathematically, if this bias is introduced by means of an exponential term, such a distribution is called a *Gibbs measure*, and the function that encodes this biasing is called a *Hamiltonian*. For instance, for a simple graph G on n labeled vertices with $E(G)$ edges and $T(G)$ triangles, we define the Hamiltonian

$$\mathcal{H}_{n;\alpha,h}(G) := \frac{\alpha}{n} T(G) + h E(G), \quad \text{with } \alpha, h \in \mathbb{R}. \tag{1}$$

As a probability measure on the space $\mathcal{G}_n$ of simple graphs with n vertices, we take the following

Gibbs probability

$$\mu_{n;\alpha,h}(G) := \frac{\exp\left(\mathcal{H}_{n;\alpha,h}(G)\right)}{Z_{n;\alpha,h}}, \quad \text{with} \quad Z_{n;\alpha,h} := \sum_{G\in\mathcal{G}_n} \exp\left(\mathcal{H}_{n;\alpha,h}(G)\right), \tag{2}$$

where the normalizing constant $Z_{n;\alpha,h}$ is called *partition function*. Random graphs whose distribution is a Gibbs measure of the form (2) are called *exponential random graphs*. When the Hamiltonian is of the form (1), we speak of the *edge-triangle* model; the well-known *Erdős–Rényi random graph* is given by the special case $\alpha = 0$ and $h = \log\frac{p}{1-p}$. A crucial characteristic of the model is the so-called *limiting free energy* associated with (2), which is defined as

$$f_{\alpha,h} := \lim_{n\to+\infty} \frac{1}{n^2} \ln Z_{n;\alpha,h}.$$

A lack of analyticity in this function characterizes the presence of a phase transition. An explicit expression of $f_{\alpha,h}$ has been obtained in Chatterjee and Diaconis, 2013, when the parameters (α, h) lie in a specific region called the *replica-symmetric regime*, corresponding to $\alpha > -2$, $h \in \mathbb{R}$. This term is borrowed from *spin glass* theory and is related to the fact that, in the limit, the model behaves like a mean-field model. In particular, it has been proved (Radin and Yin, 2013) that the replica symmetric region includes a (non-explicit) continuous and strictly decreasing curve $\mathcal{M}^{\mathrm{rs}}$ at which a first-order phase transition in the limiting edge density $u^*(\alpha, h)$ occurs, and the first-order partial derivatives of $f_{\alpha,h}$ have jump discontinuities. At the critical point $(\alpha_c, h_c) = \left(27/8, \log 2 - 3/2\right)$, the phase transition is of second order, and the second-order partial derivatives of $f_{\alpha,h}$ diverge (see Figure 9 for a qualitative representation of the phase diagram).

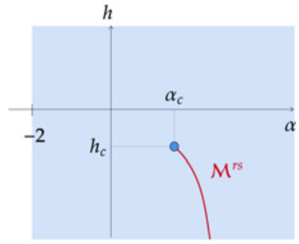

Fig. 9: *Illustration of the phase space (α, h) for the edge-triangle model (1) in the replica-symmetric regime*

One of the key results of our paper [1] is the determination of the asymptotic distribution of the edge density $\frac{2E(G)}{n^2}$ (as the graph size n tends to infinity) within the replica-symmetric regime. Our analysis provides a *strong law of large numbers* whenever the parameters (α, h) are taken outside the critical curve and proves that the edge density concentrates with high probability in a neighborhood of the *free energy maximizers* on the critical curve. Fluctuations of the edge density are also investigated, and a *central limit theorem* is derived for parameters outside the critical curve and away from the critical point (α_c, h_c). These results are extended to a general family of exponential random graphs where the Hamiltonian involves various sub-graphs counts. A predominant part of our results includes the exploration of a simplified model, the *mean-field* approximation of the edge-triangle model. A major advantage of this approximation is that the Hamiltonian can be expressed as a function of the edge density, and exact computations are possible (like in the *Curie–Weiss model*). In this setting, we can prove the analogous of the results derived for the edge-triangle model (partially in a stronger form), and we can go further, in particular, we are able to characterize the fluctuations at the critical point, presenting a *non-standard central limit theorem* with scaling exponent $3/2$.

Some heuristic computations based on *large deviation* estimates suggest that the edge-triangle model may exhibit the same behavior as the mean-field approximation when the parameters vary in the phase space. We then formulate conjectures about fluctuations at the critical point and about the behavior of the edge-triangle model on the critical curve.

Conclusion and outlook

Our work deals with phase transitions arising from random graphs occurring in diverse contexts. The models we consider reflect many real-world properties, such as clustering and being scale free, and often exhibit important features associated with the networks: the distribution of edges amongst nodes of certain weights, crucial connectivity properties, or the edge density. We therefore believe that this is an important, rich area with interesting problems both in the context of new applications, and mathematically. There are many more results in the pipeline!

References

[1] A. Bianchi, F. Collet, E. Magnanini, *Limit theorems for exponential random graphs*, arxiv Preprint no. 2105.06312, 2021.

[2] P. Gracar, L. Lüchtrath, C. Mönch, *The emergence of a giant component in one-dimensional inhomogeneous networks with long-range effects*, in: 18th International Workshop on Algorithms and Models for the Web-Graph, Toronto, Canada, May 23–26, 2023, M. Dewar, P. Prałat, P. Szufel, F. Théberge, M. Wrzosek, eds., vol. 13894 of Lecture Notes in Computer Science, Springer, Cham, 2023, pp. 19–35.

[3] T. Iyer, *Degree distributions in recursive trees with fitnesses*, Adv. Appl. Probab., **55** (2023), pp. 407–443.

[4] T. Iyer, B. Lodewijks, *On the structure of genealogical trees associated with explosive Crump–Mode–Jagers branching processes*, arxiv Preprint no. 2311.14664, 2023.

[5] T. Iyer, *On explosion in Crump–Mode–Jagers branching processes and some applications*, manuscript in preparation.

[6] B. Jahnel, L. Lüchtrath, *Existence of subcritical percolation phases for generalised weight-dependent random connection models*, arxiv Preprint no. 2302.05396, 2023.