Models of Complex Networks

Yidong Chen

May 12, 2018

Abstract In this essay, we will discuss various models of real world networks, including random graph theory (Erdos-Renyi), percolation theory, small-world networks (Watts-Strogatz), and scale-free networks.

Chapter 1

Random Graph Theory

1.1 Introduction

The first model of complex networks was given by mathematicians Paul Erdos and Alfred Renyi, who introduced probablistic methods to study graph theory. Although random graph models do not faithfully represent some key properties of real world networks, random graph theory, due to its mathematical rigor, still serves as a benchmark for theoretical study of complex networks. Many concepts first studied in random graph theory (degree distribution, clustering coefficient, characteristic path length, etc.) are now indispensible in studying complex networks. In this chapter, we will introduce the basics of random graph theory. Note that although proofs will not given for the lack of space, results in this chapter can all be proven rigorously.

1.2 Definition

We will start with some basic definitions. Mathematically, a graph is defined by the following set of data: a (finite) set of vertices V and a (finite) set of edges E, each edge connecting exactly two of the vertices. Certainly we can study infinite graph, but for our purpose of modeling real world netwokrs, it is hardly appropriate to use infinite graph. Note that in this essay we will focus on undirected graphs. There are important examples (e.g. world wide web) that are more naturally described by directed graph, but the properties that we will study in this essay can all be faithfully represented by undirected graphs. A complete graph with n vertices is a graph where any pair of vertices is connected. A subgraph is a subset of another graph. For a given graph G, we will denote the number of vertices as v(G) and the number of edges as e(G).

Now equipped with the appropriate language, we can define a random graph. There are in fact three equivalent ways to describe a random graph. For our purpose, we shall focus on the following definition. Let S(n) be the set of all subgraphs of the complete graph with n vertices. A random graph G is a random variable which takes value in S(n) with probability $Prob(G = G_0) = p^{e(G_0)}(1-p)^{\frac{n(n-1)}{2}-e(G_0)} \ \forall G_0 \in S(n)$. In other words, a random graph with n vertices is a subgraph of the complete graph with n vertices, where each edge exists with probability p. Note we need two parameters to describe a random graph: number of vertices n and edge probability p.

1.3 Observables of Random Graph

Although a highly abstract model of real world network, a random graph still contains vast amount of information. In a more physical language, there are numerous observables we can compute from a random graph. Roughly speaking, the observables are of two classes: local observables and global observables. Local observables reveal structure in a neighborhood of a given vertex, while global ones reveal topological structure of a graph.

1.3.1 Local Observables

An easy characterization of local structure is degree distribution. Degree of a vertex is defined as the number of edges connected to that vertex. Let X_k be the number of vertices with degree k (this is a random variable), then the degree distribution of a random graph is defined as the probability distribution of X_k. When the number of vertex approaches infinity, the asymptotic form of the degree distribution approaches Poisson distribution $P(X_k = r) = \exp^{-\lambda_k} \frac{\lambda_k^r}{r!}$, where the mean $\lambda_k = NC_{N-1}^k p^k (1-p)^{N-1-k}$. The exponential tail of this Poisson distribution indicates that in any random graph, the number of vertex with degree k does not deviate too much from the mean value. Unfortunately, this is not the general behavior of a real world network. Instead of an exponential tail, many networks (e.g. world wide web) have degree distribution that decays as power law. Such networks are now called scale-free networks, and we will talk about them later in the essay.

Another useful local observable is clustering coefficient. Given the immediate neighbors of a vertex, clustering coefficient describes how closely related these vertices are. More precisely, consider vertex v, let k_v be the degree of v and let E_v be the number of edges that exist between these k_v vertices, then clustering coefficient of vertex v is $C_v = \frac{2E_v}{k_v(k_v-1)}$. Since C_v is a random variable depending on v, we will define clustering coefficient C

of the random graph as the mean of C_v. For a random graph with edge probability p, $C = p$ as the number of vertices approaches infinity. If we compare a real network to a random graph with the same number of vertices and same average number of edges per vertex, the clustering coefficient of the real network is usually higher than that of the random graph. Such high clustering coefficient is typical of a regular lattice, and indicates that real networks have well defined local neighborhoods.

1.3.2 Global Observables

Global observables are those quantities that do not change under local morphisms of a graph. One global observable is the characteristic path length of a random graph, which describes the average distance between two vertices. Given two vertices v and w from a random graph, suppose they are connected, then the distance between v and w is defines as the number of edges in the shortest path between v to w. (If v and w are disconnected, then the distance is defined to be infinite.) And the characteristic path length is defined as the mean distance averaged over all pairs of connected vertices. A theorem proved by Lu and Chung shows that, as the number of vertices approaches infinity, the characteristic path length almost surely approaches $\frac{\log(n)}{\log(np)}$, where n is the number of vertices and p is the edge probability. Numerical calculation shows that this quantity is usually small (of order unity) for a random graph, which shows that globally a random graph is closely related since it does not take long to go from one vertex to another.

A more fundamental topological property of a random graph is its connectedness. If a graph is connected, then there are no isolated vertex. Whether isolated vertex exists or not depends on the two parameters n (number of vertices) and p (edge probability). If p=1, we have a complete graph and no isolated vertex exists. If p=0, all vertices are isolated. Therefore, for a fixed n, there must be a threshold value of p above which isolated vertices disappear. This is our first example of a phase transition in a random graph. More precisely, we can prove that for a fixed n, this threshold is given by $\frac{\log(n)}{n}$. Furthermore, we can show that this threshold is sharp so that $\forall p < \frac{\log(n)}{n}$, isolated vertices almost surely exist, and $\forall p > \frac{\log(n)}{n}$, isolated vertices almost surely disappear.

In fact, many properties of random graphs undergo phase transition. One prominent example is the sudden emergence of a giant component. If we study the connected components of large networks (be it random graphs or real networks), there is often one component that's much larger than the other components. For example, in the protein-protein interaction network in E. Coli, there is only one connected component containing 1851 vertices

while all the other connected components contain fewer than 16 vertices. In the context of random graph, a giant component is defined to be a connected component where the number of vertices contained is proportional to n (the total number of vertices). We can think of the giant component as a component that spans across the entire graph. This is certainly a stronger notion of connectedness than the disappearance of isolated vertex. It is proven that $\forall p = \frac{d}{n}$, where $d > 1$, a giant connected componnet almost surely exits. This is a seminal result in random graph theory, and it captures a key property of real world networks.

Chapter 2

Percolation Theory

2.1 Introduction

Percolation theory was first introduced as a way to model the flow of fluid through porous medium of small chanels which may or may not let the fluid pass. It is the simplest not exactly solvable model with a phase transition. As such, percolation theory provides valuable insights into more complicated physical systems. Outside of physics, percolation theory is important in studying biological systems, geophysics, and network science. In the context of network science, percolation theory provides an alternative to random graph theory. We will introduce the basics of percolation theory in this chapter and see how it helps us understand complex networks from a different perspective.

There are several versions of percolation theory; for our purpose, we will focus on bond percolation. Consider a d-dimensional regular lattice where the edge exists with probability p. The main problem bond percolation theory addressese is the emergence of path that percolate the lattice. There is a critical percolation threshold p_c such that $\forall p > p_c$ an infinite percolating cluster almost surely exists. This cluster spans across the entire lattice, and as the lattice size approaches infinity, the size of this cluster diverges (hence the name infinite cluster).

A key observable in percolation theory is the percolation probability P_s, which is the probability that the cluster at the origin has size s. (The size of a cluster is the number of vertices it contains. And due to regularity of the lattice, there's no loss of generality by focusing on the cluster at the origin.) For $p < p_c$ (subcritical phase), only small clusters exist. To measure the size of these clusters, we consider the rate of decay of P_s as s approaches infinity. In general, it can be shown that the asymptotic form of P_s is given by $\exp^{-\frac{r}{\xi}}$, where r is the radius of the cluster and ξ should be treated as

correlation length, which will diverge as p approaches p_c from below. This exponential decay indicates that the mean radius of the finite cluster is given by ξ, which diverges as the order parameter p approaches the critical value.

For $p > p_c$, the correlation length ξ is no longer useful. Instead, we consider the probability P that the origin is contained in the infinite cluster. Near the critical point, the scaling form of P with respect to $p - p_c$ contains information about the phase transition (critical exponent).

2.2 Percolation on Cayley Tree

In this section, we will consider percolation on Cayley tree. This is an exactly solvable problem, and is closely related to infinite dimensional percolation which we will talk about in the next section.

A Cayley tree is a tree (a graph without cycles) where each vertex has z neighbors, except vertices at the surface. The unique property of a Cayley tree is that its number of vertices at the surface (surface area) is proportional to the total number of vertices (volume). In a usual regular lattice (e.g. honeycomb, hypercubic, etc.), surface area is proportional to volume$^{1-\frac{1}{d}}$, where d is the dimensionality. Only when d goes to infinity is the surface area proportional to volume. In this sense, Cayley tree represents an infinite dimensional object.

If we consider percolation on Cayley tree with coordination number z and percolation probability p, then it can be shown that percolation threshold $p_c = \frac{1}{z-1}$. For p close to but larger than p_c, the percolation probability P that the origin is contained in the infinite cluster obeys the following scaling law: $P \propto (p - p_c)$. For p close to but smaller than p_c, the correlation length ξ obeys the following scaling law: $\xi \propto (p_c - p)^{-\frac{1}{2}}$.

2.3 Infinite Dimensional Percolation

Now consider percolation on a d-dimensional regular lattice, where d is larger than the upper critical dimension. In this case, phase transition is faithfully descirbed by mean field results, which state: As $p \to p_c$, the percolation probability $P \propto |p - p_c|$ and the correlation length $\xi \propto |p - p_c|^{-\frac{1}{2}}$. Certainly, these results hold when d goes to infinity. Thus, we see that Cayley graph does faithfully reproduce infinite dimensional results.

It's interesting to ask what the upper critical dimension is. Although it is widely believed to be 6, a rigorous result proved by Hara and Slade can

only pin down d_c to be 19. Nevertheless, for our purpose which is to study the limit where d goes to infinity, the actual value of d_c does not concern us.

2.4 Percolation and Random Graph

To see why we care about infinite dimensional percolation, let's consider how we can embed a random graph with n vertices into a regular lattice. Since any pair of vertices can be connected in a random graph, the coordination number will have to be $N-1$ to make the embedding work. Once we have the embedding, we can interpret the edge probability of the random graph as the percolation probability of the regular lattice. Now, since in random graph theory we are concerned with asymptotic behavior when n goes to infinity, it is equivalent to studying infinite dimensional percolation.

As concrete evidence of this equivalence, let's compare percolation threshold to the critical probability where giant cluster first appears in a random graph. Since infinite dimensional percolation can be described by percolation on a Cayley tree, percolation threshold $p_c = \frac{1}{z-1}$. Now to embed a random graph with n vertices into a Cayley tree, we need $z = n - 1$. Thus $p_c = \frac{1}{n-2} \approx \frac{1}{n}$, when n goes to infinity. Yet, $\frac{1}{n}$ is exactly the critical probability when giant cluster first appears in a random graph with n vertices.

In fact, the correspondence between infinite dimensional percolation and random graph shows that phase transition in both contexts belongs to the same universality class. This is confirmed by numerical simulations of random graphs (Christensen et al., 1998). Despite the equivalence, it is often useful to study a complex network from both perspectives (random graph and percolation). While random graph theory more naturally addresses question like the disappearance of isolated vertex, percolation theory provides a more natural way to study phase transition and cluster distribution.

Chapter 3

Small-world Network

3.1 Introduction

The previous two chapters introduced two classical ways to study complex networks. While percolation theory assumes the existence of an underlying regular lattice, random graph theory does not make this assumption and is equivalent to infinite dimensional percolation. However, many real world networks cannot be faithfully represented by either approach. As discussed in chapter 1, a random graph describes a network with a small clustering coefficient (equals to the edge probability p) and a small characteristic path length (roughly goes as $\log n$, where n is the number of vertices). On the other hand, percolation on regular lattice describes a network with a large clustering coefficient (as coordination number $z \to \infty$, clustering coefficient approaches $\frac{3}{4}$) and a large characteristic path length (roughly goes as $n^{\frac{1}{d}}$, where d is the dimensionality and n is the size of the lattice). Yet, numerical studies show that real world networks usually have a large clustering coefficient and a small characteristic path length. This character is called small-world property. To model such small-world networks, we need a way to interpolate between regular lattice and random graph.

3.2 Watts-Strogatz Model

Watt and Strogatz made the following key observation. Since real networks usually have large clustering coefficients, the local behavior of a real network should be similar to that of a regular lattice. Now to decrease characteristic path length of a regular lattice, we need to connect vertices far apart with short cuts. Notice that, as a model of real network, the number of edges and the number of vertices are model inputs. Thus we cannot simply add edges to connect vertices far apart. The best we can do is to eliminate some edges connecting neighboring vertices and rewire it to connect vertices far apart. Watt and Strogatz realized that we only need a small number

of such rewired edges to reproduce a small characteristic path length while retaining a large clustering coefficient. One rewired edge between A and B decreases distance between points in the neighborhood of A and points in the neighborhood of B. Thus each rewired edge has a very pronounced effect in cutting characteristic path length.

With this observation, we now present the original Watts-Strogatz construction. There are numerous modifications, but to illustrate the main idea we will focus on this original construction. Start with a regular ring lattice with n vertices. Assume the coordination number is z, meaning each vertex is connceted to its immediate z neighbors (z/2 on either side). Now introduce disorder into the lattice by rewiring each edge with probability p. Rewiring means moving one end of an edge to a new randomly selected vertex. Self-connections and duplicate edges will be excluded. The rewiring probability p gives a one-parameter interpolation between regular lattice and random graph. For p=0, we have a regular lattice; for p=1, we have a random graph. And for small p, we should expect to have a model of small-world network.

3.3 Properties of Small-world Network

We first study how characteristic path length $l(n, z, p)$ scales as rewiring probability p. (n, z, p are the three parameters in Watts-Strogatz model, so we should expect l to depend on all three parameters.) It is clear that as p goes from 0 to 1, there is a qualitative change of behavior in l. This suggests that a renormalization group analysis should help reveal the scaling form of l. Indeed, a study by Newman et al. shows the following scaling form: $l(n, z, p) \approx \frac{n}{z} f(pzn^d)$, where d is the dimensional of the original regular lattice and f is a universal function that obeys: $f(x)$ =const, for $x << 1$; $f(x) = \frac{\log(x)}{x}$, for $x >> 1$. Thus, to have a characteristic path length similar to a random graph (which scales as $\frac{\log(n)}{n}$), we only need rewiring probability p to be larger than the critical value $\frac{1}{zn^d}$. This value is very small when the number of vertices n goes to infinity.

Similarly we can ask for the scaling form of clustering coefficient $C(n, z, p)$. Previously we defined clustering coefficient as $C =< \frac{2E_v}{k_v(k_v-1)} >$, where E_v is the number of edges between the neighbors of a vertex, k_v is the degree of vertex v and the bracket means taking the average. Following the convention in literature, here we adopt a slightly different notion of clustering coefficient, denoted C'. $C' = \frac{<E_v>}{<\frac{k_v(k_v-1)}{2}>}$. Now it is shown that the scaling form of $C'(n, z, p)$ at small p is $C'(n, z, p) \approx C'(0)(1-p)^3$, where $C'(0)$ is the clustering coefficient of a regular lattice. Thus for small p, we see

that the clustering coefficient is roughly that of a regular lattice. Therefore, for small rewiring probability p, Watts-Strogatz model successfully produces a small characteristic path length and a large clustering coefficient.

We can also study the degree distribution of Watts-Strogatz model. At p=0, we have a regular lattice. Thus the degree distribution is a delta function with a peak at the coordination number z. As p increases, disorder broadens the delta distribution but retains a central peak at z. It is shown by Barrat and Weigt that the full functional form of the degree distribution is a modified Poisson distribution with an exponential tail. This is similar to random graph model.

Chapter 4

Scale-free Network

As we've seen above, both random graph and Watts-Strogatz model has a degree distribution with an exponential tail. However, there are real networks whose degree distributions have a power-law dependence. Such networks are called scale-free. Most prominent example of scale-free network is the world wide web. The construction of a scale-free network is significantly different from constructing a random graph or a Watts-Strogatz model. Thus, to keep the paper brief, we shall not go into the details of constructing a scale-free network.

One key assumption of random graph and Watts-Strogatz model is that the number of vertices is fixed. In reality, this may not be the most natural way to describe a network. For example, the world wide web grows exponentially as new webpages (vertices) are added. In addition, many networks exhibit preferential attachment: the probability of connecting a vertex depends on the degree of that vertex. For example, a new webpage is more likely to connect to a popular webpage with a large number of existing connections (a large degree).

These two features - growth and preferential attachment - are captured by scale-free models. Start with a small number of vertices, at each timestep scale-free model adds a new vertex with a given number of edges that connect to existing vertices in the network (growth). The probability that a new vertex will connect to an existing vertex v depends on the degree k_v (preferential attachment). From this construction, we can compute the degree distribution of the network. It turns out the result is a power-law distribution.

Bibliography

[1] Reka Albert, and Albert-Laszlo Barabasi. *Statistical Mechanics of Complex Networks*. arXiv:cond-mat/0106096v1.

[2] Duncan J. Watts, and Steven H. Strogatz. *Collective Dynamics of 'Small-world' Networks*. Nature. 393, 6684, 440, June 4, 1998.

[3] Dietrich Stauffer, and Amnon Aharony. *Introduction to Percolation Theory (Revised Second Edition)*. Taylor and Francis, 1994.

[4] A. Barrat, and M. Weigt. *On the Properties of Small-world Network Models*. Eur. Phys. J. B 13, 547-560, 2000.

[5] M. E. J. Newman, and D. J. Watts. *Renormalization Group Analysis of the Small-world Network Model* Physics Letters, Section A: 263, 341-346, 1999.

[6] M. Barthelemy, and L. Nunes Amaral. *Small-world Networs: Evidence for a Crossover Picture* Princeton University Press, 304-307, Oct. 23, 2011.

[7] Edward M. Marcotte, Matteo, Pellegrini, Ho-Leung Ng, Danny W. Rice, Todd O. Yeates, and David Eisenberg. *Detecting Protein Function and Protein-Protein Interactions from Genome Sequences* Science. 5428, 751, 1999.

[8] Takashi Hara, and Gordon Slade. *Mean-field Critical Behaviour for Percolation in High Dimensions* Commun. Math. Phys. 128, 333-391, 1990.