Phase Transitions in Complex Networks

Thomas Tuegel

May 11, 2012

Abstract

Recent technological developments have enabled the creation and study of complex networks in information technology, biology, and other areas, renewing interest in the study of this field. The basic theory of networks is outlined, with the goal of describing the structure of various networks. Phase transitions in network structures themselves, such as the emergence of the giant connected component and the condensation of edges, are described. Methods used to study simple models on networks are described. The resulting descriptions of phase transitions and critical behavior in the Ising model and simple epidemiological models are reported.

1 Introduction

Complex networks are ubiquitous in the natural and man-made world. This understanding, however, is tied to the technological developments of the last 60 years. For example, it is now known through genetic sequencing that the metabolic pathways of many organisms share a common network structure [8]. It has also become possible not only to simulate large networks computationally, but construct them for practical use: the Internet and World Wide Web are observed to have complex network structures [7, 5], while the latter is believed to have a size in excess of 10^{11} pages [5]. These developments have energized this field of study, some of the results of which are described below.

2 Characterization of Networks

2.1 Definitions from Graph Theory

Mathematically, a network is understood as a graph: a set of *vertices* and a set of *edges*. An edge is an unordered pair of vertices, its endpoints, which are said to be *adjacent* by virtue of the existence of the edge. An edge is a *loop* if its endpoints are identical. Two edges are *parallel* if the share the same pair of endpoints. [9] *Simple* graphs have neither loops nor parallel edges [9]; the study of networks concerns itself primarily with such graphs [5].

The number of edges connected to a vertex is its *degree*, conventionally denoted q [9]. Naturally, one can construct the degree distribution of a network, $P(q)$, the probability that a vertex has degree q. Though the degree distribution is important to the characterization of networks, alone it is generally insufficient to completely describe the architecture of a network. Uncorrelated networks, described below, are an exception to this rule in that they are completely described by their degree distribution. [5]

A *path* is an ordered list of vertices where the vertices are all distinct and each pair of vertices adjacent in the list is also adjacent in the graph. The length of the path is the number of edges thus defined. A *cycle* is a path with the additional constraint that the first and last vertices are adjacent. A graph which contains no cycles, but where a path can be constructed between any two vertices, is called a *tree*. [9] Some calculations on networks can be performed analytically if the network is a tree, or at least locally tree-like [5].

The *intervertex distance,* l_{ij}, is the length of the shortest path between vertices i and j. The probability distribution of intervertex distances helps characterize the overall structure of a network, while the compactness of the network is characterized by the mean intervertex distance, $\bar{l}$. The *diameter* of the network is the largest intervertex distance between any two vertices. [5]

A graph can also be characterized by its *load distribution,* $P(b)$. The *load* of a vertex is related to the number of shortest paths passing through that vertex. If $s(i, j)$ is the number of shortest paths between vertices i and j, and $s(i, k, j)$ is the number of shortest paths between vertices i and j passing through k, then the load is given by [5]:

$$b(k) = \sum_{i \neq j,\, i,j \neq k} \frac{s(i, k, j)}{s(i, j)} \tag{1}$$

The *neighborhood* of a vertex is the set of all vertices adjacent to it; i.e., $u \in N(v)$ if and only if (u, v) is an edge in the graph. A *clique* is a graph or a subgraph with every possible edge, given its set of vertices. In a clique, the neighborhood of any vertex is the entire graph. [9] A *k-clique* is a clique with k vertices; a *triangle* is a common name for a 3-clique [5]. A vertex can be characterized by its *clustering coefficient*,

$$C(v) = \frac{2t_v}{q_v(q_v - 1)} \tag{2}$$

where t_v is the number of triangles containing the vertex and q_v is the degree of the vertex. The maximum number of possible triangles through a vertex is $\frac{1}{2}q_v(q_v - 1)$, so that the clustering coefficient is merely the ratio of actual triangles to possible triangles. The mean clustering coefficient, $\langle C \rangle$, is useful in characterizing networks. [5]

2.2 Regular Trees

A *regular* tree of degree q has vertices which are all either of degree q or degree 1. A *Bethe lattice* is an infinite, regular tree. Note that all vertices in a Bethe lattice are equivalent because there are no edges. A *Cayley tree* is a finite regular tree with a central vertex; it can be viewed in some sense as a Bethe lattice which has been truncated at an arbitrary distance from the central vertex. The behavior of simple models is often analyzed on these networks because their regularity eases analytic calculations and the results are often universal across locally tree-like networks. [5]

2.3 Random and Uncorrelated Networks

A random network simply has a random assortment of edges. Specific types of random networks are defined by the subsequent restraints placed on randomness. One general class of random networks are the uncorrelated networks. Uncorrelated networks have no correlation between the degrees of neighboring vertices. In a general random network,

$$P_{N(v)}(q) = \frac{qP(q)}{\langle q \rangle} \tag{3}$$

$$\langle q \rangle_{N(v)} = \frac{\langle q^2 \rangle}{\langle q \rangle} \tag{4}$$

where $P_{N(v)}$ is the degree distribution and $\langle q \rangle_{N(v)}$ is the mean degree in the neighborhood of vertex v [5]. In an uncorrelated network, however, there are no such correlations and

$$P_{N(v)}(q) = P(q) \tag{5}$$

$$\langle q \rangle_{N(v)} = \langle q \rangle \tag{6}$$

Because of this fact, uncorrelated networks are described completely by their degree distributions. [5]

Dorogovtsev et al. describe the Erdős-Rényi model as an ensemble of simple, uncorrelated networks. In this ensemble, all the possible graphs with a fixed number of vertices and edges

have equal probability. For graphs with N vertices and M edges, the mean degree is

$$\langle q \rangle = \frac{2M}{N} \tag{7}$$

The Gilbert model is also an ensemble of graphs with a fixed number of vertices, but rather than have a fixed number of edges, each of the possible edges appears with a fixed probability p. The mean degree of a graph in the Gilbert model is

$$\langle q \rangle = pN \tag{8}$$

Both models share a Poisson degree distribution

$$P(q) = \frac{e^{-\langle q \rangle}}{q!} \langle q \rangle^q \tag{9}$$

For this reason, the Erdős-Rényi and Gilbert models are collectively known as classical random graphs. [5]

Complex networks are defined as those networks that have degree distributions other than the Poisson distribution of the classical random graphs. Equilibrium random trees are ensembles of complex networks where each member is a random, connected tree with a particular number of randomly labeled vertices. If the trees are equally weighted, the degree distribution is

$$P(q) = \frac{e^{-1}}{(q-1)!} \tag{10}$$

[5]. If particular weights are introduced [2], *scale-free* networks can be produced; these have the degree distribution

$$P(q) \propto q^{-\gamma} \tag{11}$$

and are of interest to the study of critical phenomena [5].

3 Structural Phase Transitions

3.1 Giant Connected Component

The giant connected component is an extensive component which emerges in certain networks in the sparse ($\lim_{N\to\infty} \frac{\langle q \rangle}{N} = 0$) limit. The size of the giant connected component can be determined from

$$1 - S = \sum_q P(q) x^q \tag{12}$$

where S is the fraction of vertices in the giant connected component and

$$x = \sum_q \frac{qP(q)}{\langle q \rangle} x^{q-1} \tag{13}$$

The giant connected component exists if $S > 0$ in the thermodynamic limit. The condition that the giant connected component exist is the existence of some $x < 1$ which solves Equations 12 and 13. It is found that in uncorrelated networks, the giant connected component exists if

$$\left\langle q^2 \right\rangle - \langle q \rangle > \langle q \rangle$$

where $\langle q^2 \rangle - \langle q \rangle$ is the average number of next-nearest neighbors of a vertex. Note that this inequality is automatically satisfied if $\langle q^2 \rangle \to \infty$, which is the case for scale-free graphs with $\gamma \leq 3$. [5]

The robustness of the giant connected component in uncorrelated is studied using percolation methods, where a fraction $1 - p$ of vertices or edges are removed randomly from the network. It is found that under this process, the giant connected component remains intact when

$$p\left(\left\langle q^2 \right\rangle - \langle q \rangle\right) > \langle q \rangle \tag{14}$$

$$p > p_c = \frac{\langle q \rangle}{\langle q^2 \rangle - \langle q \rangle} \tag{15}$$

Now we see that not only is the giant connected component guaranteed to exist if $\langle q^2 \rangle \to \infty$, but in this limit $p_c \to 0$ so that the giant connected component cannot be destroyed by randomly removing a fraction of its vertices or edges! [5]

The remarkable robustness of certain scale-free networks against random damage has important implications for the design of real-world networks. For example, a study of the global architecture of the Internet revealed a scale-free degree distribution with $\gamma \approx 2.17$, indicating its overall resistance to random failure [7, 1]. Biological systems are also noted for their resistance to random failure [1]. Jeong et al. undertook a study of the metabolic pathways of 43 organisms, modeling them as networks where vertices represent substrates (molecules involved in the metabolic process) and edges represent reactions [8]. They found that these networks are scale-free with exponent γ ranging from 2.0 to 2.4 [8]. Although their study was limited by the availability of genetic sequences and even included organisms for which only a partial sequence was available, their results are nevertheless suggestive. These examples highlight the incredible robustness of certain types of scale free networks, but it must also be noted that these network architectures rely on the presence of a few highly-connected (high degree) vertices. Removing some of these vertices quickly destroys the giant connected component of these networks, rendering them highly susceptible to directed (rather than random) damage or attack. [5, 1]

3.2 Condensation

A condensation transition refers to the "condensation" of some extensive number of some network structure (edges, triangles, etc.) into a subnetwork much smaller than the network itself. (Smaller in the sense of having a smaller diameter.) [5] This transition takes place as some random network is evolved toward equilibrium according to some chosen free energy [3] or preference function [6]. For example, a network may be evolved with a degree dependent preference function $f(q)$ where at each step, an edge may created randomly between two vertices i and j with probability $f(q_i) f(q_j)$ according to the degrees of the vertices [6, 5]. Choosing a preference function such as

$$f(q) \sim q + 1 - \gamma + \mathcal{O}\left(q^{-1}\right) \tag{16}$$

produces scale-free networks with degree distribution $P(q) \propto q^{-\gamma}$ at the critical value $\langle q \rangle = q_c$. Above q_c, a fraction of all the edges in the network attach to a single vertex, an example of

an edge condensation transition. (The degrees of the remaining vertices are still distributed according to the scale-free distribution.) [6, 5]

Another approach to the study of the condensation transition is to define an appropriate temperature parameter and a free energy and evolve the starting graph to equilibrium according to statistical mechanics. Derényi et al. take this approach, using Monte Carlo simulation to evolve a network according to one of several choices for the free energy. They find a variety of phase transitions in the network structure, including a second order transition from the disordered state into an edge-condensed phase and, for some values of $\Phi_q = \frac{q_{max}}{M}$, a subsequent first order transition from the edge-condensed phase into a fully connected phase (here, M is the number of edges on the graph). [3]

4 Ising Model

The Ising model is likely familiar to most physicists, even outside the field of complex networks. The Hamiltonian which describes the Ising model on a network is only a slight variation from the more familiar lattice formulation:

$$\mathcal{H} = -\sum_{i<j} J_{ij} a_{ij} S_i S_j - \sum_i H_i S_i \tag{17}$$

where a_{ij} is the adjacency matrix: $a_{ij} = 1$ if there exists a path connecting two vertices or 0 otherwise [5]. Complex networks are infinite-dimensional in the thermodynamic limit, so one would expect mean field theory to be valid (the upper critical dimension of the Ising model being $d_c = 4$). Adopting a simple nearest-neighbor Weiss mean field approach,

$$\mathcal{H} = -\sum_i H_i S_i \tag{18}$$

where $H_i = H + J \langle q \rangle M$. There is a uniform external field, H, with the average magnetization denoted by M. This leads to the usual equation for the magnetization

$$M = \tanh\left(\beta H + \beta J \langle q \rangle M\right) \tag{19}$$

This predicts a critical temperature given by $T_c = J \langle q \rangle$ with magnetization and susceptibility given respectively by

$$M \sim t^{\beta} \tag{20}$$

$$\chi \sim |t|^{-\tilde{\gamma}} \tag{21}$$

with $\beta = \frac{1}{2}$ and $\tilde{\gamma} = 1$. This is the expected mean field behavior for the Ising model, and early computer simulations of complex networks agreed with these predictions. However, subsequent simulation showed that in some scale-free networks, the critical temperature diverges as $T_c \sim \log N$. [5]

The true behavior of the ferromagnetic Ising model on an uncorrelated, complex network can be determined using the Bethe-Peierls approximation. Assume a uniform external field

H and an interaction $J_{ij} = J > 0$. Consider the neighborhood of vertex i; the Hamiltonian for this neighborhood can be written

$$\mathcal{H} = -\sum_{j \in N(i)} JS_iS_j - HS_i - \sum_{j \in N(i)} \phi_{j\backslash i} S_j \tag{22}$$

where the additional field $\phi_{j\backslash i}$ has been introduced to account for the interactions within the neighborhood that are not explicitly handled in the Hamiltonian [5]. The total field at vertex i is

$$H_i^{tot} = H + \sum_{j \in N(i)} h_{ji} \tag{23}$$

where h_{ji} is the field at vertex i due to the spin at vertex j, given by

$$\tanh \beta h_{ji} = \tanh \beta J \tanh \beta \phi_{j\backslash i} \tag{24}$$

[5]. The total field at vertex j is

$$H_j^{tot} = \phi_{j\backslash i} + h_{ij} \tag{25}$$

where

$$\tanh \beta h_{ij} = \tanh \beta J \tanh \left[\beta \left(H + \sum_{n \in N(i), n \neq j} h_{ni} \right) \right] \tag{26}$$

[5]. By considering the same equations around a neighborhood of vertex j instead, one finds that

$$\phi_{i\backslash j} = H + \sum_{n \in N(i), n \neq j} h_{ni} \tag{27}$$

On an uncorrelated network, the fields h_{ij} should be random with a distribution given by

$$\psi(h) = \sum_q \frac{P(q)\, q}{z_1} \int \delta \left(h - T \tanh^{-1} \left\{ \tanh \beta J \tanh \left(\beta H + \beta \sum_{m=1}^{q-1} h_m \right) \right\} \right) \prod_{m=1}^{q-1} \psi(h_m)\, dh_m \tag{28}$$

so that the magnetization is

$$M = \sum_q P(q) \int \tanh \left(\beta H + \beta \sum_{m=1}^{q} h_m \right) \prod_{m=1}^{q} \psi(h_m)\, dh_m \tag{29}$$

where z_1 is the average number of nearest neighbors in the network [5]. The critical temperature is

$$T_c = \frac{2J}{\log \left(\frac{z_2 + z_1}{z_2 - z_1} \right)} \tag{30}$$

where z_2 is the average number of next-nearest neighbors in the network [5]. The scaling behavior of scale-free networks is described in Table 1. Observe that for scale-free networks with $\gamma > 3$, the Ising model obeys the scaling law

$$\alpha + 2\beta + \tilde{\gamma} = 2 \tag{31}$$

Furthermore, for $\gamma > 5$ the model follows the mean-field predictions for the Ising model, but for $\gamma < 5$, the critical exponents depend on γ. [5]

Table 1: Scaling behavior of the ferromagnetic Ising model on scale-free networks over a range of values of γ [5]. (Note that $t = \frac{T-T_c}{T_c}$.)

	M	C	χ
$\gamma > 5$	$t^{\frac{1}{2}}$	discontinuous	t^{-1}
$\gamma = 5$	$\left(\frac{t}{\log(t^{-1})}\right)^{\frac{1}{2}}$	$\frac{1}{\log(t^{-1})}$	t^{-1}
$3 < \gamma < 5$	$t^{\frac{1}{\gamma-3}}$	$t^{\frac{5-\gamma}{\gamma-3}}$	t^{-1}
$\gamma = 3$	$e^{-\frac{2T}{\langle q \rangle}}$	$T^2 e^{-\frac{4T}{\langle q \rangle}}$	T^{-1}
$2 < \gamma < 3$	$T^{-\frac{1}{3-\gamma}}$	$T^{-\frac{\gamma-1}{3-\gamma}}$	T^{-1}

5 Epidemiological Models

The study of networks has greatly enhanced the understanding of epidemics. Prior to the development of network models, most work on disease spreading was done under the assumption that disease carriers interact with susceptible individuals in their population at random. This assumption is obviously invalid and in fact, it is now known that in many cases, the number of contacts made by individuals in a population follows a scale-free distribution. [10]

The oldest model of disease spread is the SIR model [10]. In this model, individuals (vertices) are in one of three states: susceptible to infection, infected and able to transmit the disease to others, and removed, either due to recovery and subsequent immunity or death [10, 5]. Let $s(t)$, $i(t)$, and $r(t)$ be the fraction of individuals in the population that are susceptible, infective, and removed, respectively. Clearly,

$$s(t) + i(t) + r(t) = 1 \tag{32}$$

[10, 5]. Prior to the development of network models, the simplifying assumption was made that infected individuals interact with susceptible individuals randomly at rate β. Then, if infective individuals recover or die at rate γ, the following system of differential equations governs the spread of the disease:

$$\frac{ds}{dt} = -\beta i(t) s(t) \tag{33}$$

$$\frac{di}{dt} = \beta i(t) s(t) - \gamma i(t) \tag{34}$$

$$\frac{dr}{dt} = \gamma i(t) \tag{35}$$

[10].

Before addressing corrections to this model due to network structure, it is also worth discussing other types of epidemiological model. Dorogovtsev et al. [5] describe four basic models: If infected individuals never die or recover, a disease would be modeled using the SI model where vertices never leave the infective state. If the infected recover, but recovery grants no immunity, a disease would be modeled using the SIS model, where infected vertices return to the susceptible state at a certain rate. A generalization of the SIR model is the

SIRS model, in which the immunity granted by recovery is temporary. The discussion here will be limited to the simpler SI, SIR and SIS models.

Applying a network structure to these models is as simple as adapting the rule for infection transmission. On a network, susceptible vertices are infected randomly by their infective nearest neighbors at rate λ [10, 5]. The infection rate is obviously the only parameter of the SI model, but the SIR and SIS models can be rescaled and written entirely in terms of this one parameter [5]. The differential equations governing the spread of a disease in the SIR model on an uncorrelated network become

$$\frac{dr_q}{dt} = i_q(t) \tag{36}$$

$$\frac{di_q}{dt} = -i_q(t) + \lambda q\left[1 - i_q(t)\right]\sum_{q'}\frac{q'-1}{\langle q\rangle}P(q') \tag{37}$$

Notice that $r(t)$ and $i(t)$ have become $r_q(t)$ and $i_q(t)$, the fraction of removed and infective vertices *of degree* q. This leads to qualitatively different behavior from the non-network model. [5]

In all of these models, there is a phase transition at the epidemic threshold, λ_c. Dorogovtsev et al. [5] characterize the epidemic threshold in the following way: Below the epidemic threshold, $\lim_{t\to\infty} i(t) = \lim_{t\to\infty} r(t) = 0$ in the thermodynamic limit. In the SIS model, an epidemic is characterized by $\lim_{t\to\infty} i(t) > 0$. The extreme version of this case is the SI model, where $\lim_{t\to\infty} i(t) = 1$ for any value of λ; there is no threshold in this case. In the SIR model, it is obviously not possible to have an infection persist into the long time limit, but an epidemic is characterized by $\lim_{t\to\infty} r(t) > 0$. [5] The epidemic thresholds for uncorrelated networks are

$$\lambda_c^{SIS} = \frac{\langle q\rangle}{\langle q^2\rangle} \tag{38}$$

$$\lambda_c^{SIR} = \frac{\langle q\rangle}{\langle q^2\rangle - \langle q\rangle} \tag{39}$$

[5].

The importance of initial conditions cannot be overstated: even above the epidemic threshold, not all outbreaks will lead to epidemics. For example, scale-free networks have a few vertices with high degree, while most vertices have lower degree. Infections that start far from one of these hubs may not become epidemics, even above the epidemic threshold. [5, 10] Recent work suggests that traditional vaccination strategies may be ineffective in scale-free networks for this very reason. However, just as the giant connected component in scale-free graphs resists random damage but is exceptionally susceptible to targeted damage, strategies that target the most well connected nodes of a network for vaccination are also more effective than random vaccination. [4, 11] Targeted strategies have the added advantage of lower cost and improved cost-effectiveness. [4]

6 Conclusion

It is apparent that complex networks demonstrate a variety of phase transitions and critical behavior. The networks themselves can be shown to exhibit transitions between a variety

of ordered and unordered phases. Simple models, such as the Ising model, placed on the network are strongly influenced not only by the high dimensionality of such networks, but also by the structured phases of the network itself. The proliferation of real-world examples where network models have proven utility is sure to drive continued developments in this field.

References

[1] R. Albert, H. Jeong, and A.-L. Barabási. Error and attack tolerance of complex networks. *Nature*, 406:378–382, 2000.

[2] Z. Burda, J.D. Correia, and A. Krzywicki. Statistical ensemble of scale-free random graphs. *Phys. Rev. E*, 64, 2001.

[3] I. Derényi, I. Farkas, G. Palla, and T. Vicsek. Topological phase transitions of random networks. *Physica A*, 334:583–590, 2004.

[4] Z. Dezső and A.-L. Barabási. Halting viruses in scale-free networks. *Phys. Rev. E*, 65, 2002.

[5] S. N. Dorogovtsev, A. V. Goltzev, and J. F. F. Mendes. Critical phenomena in complex networks. *Rev. Mod. Phys.*, 80(4):1275–1335, 2008.

[6] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin. Principles of statistical mechanics of uncorrelated random networks. *Nuclear Physics B*, 666:391–416, 2003.

[7] M. Faloutsos, P. Faloutsos, and C. Faloutsos. On power-law relationships of the Internet topology. *ACM SIGCOMM Computer Communication Review*, 29(4), 1999.

[8] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A.-L. Barabási. The large-scale organization of metabolic networks. *Nature*, 407:651–654, 2000.

[9] Cheng-Kuan Lin and Lih-Hsing Hsu. *Graph Theory and Interconnection Networks*. CRC Press, 2008.

[10] M. E. J. Newman. Spread of epidemic disease on networks. *Phys. Rev. E*, 66, 2002.

[11] R. Pastor-Satorras and A. Vespignani. Immunization of complex networks. *Phys. Rev. E*, 65, 2002.