Chapter 1: Topology of Complex Networks

W. Garrett Jenkinson and John Goutsias

November 2, 2011

1 What is a Complex Network?

Complex network science involves the study of systems of relatively simple components which interact with one another in a complex fashion. The components and their interactions need not be complex individually, but the shape, or topology, of their interaction network is what puts the word “complex” in front of “network science.” After introducing some basic graph theory, we can concretely discuss what makes the topology of a network “complex.” For now, it suffices to understand that simplifying symmetries and uniform patterns are absent from the topology of a complex network. Furthermore, it is often the case that practical examples of complex networks contain an additional layer of sophistication which arises from the sheer size of the network.

Researchers who first began studying examples of large networks and properties of their topological structure were amazed to find that striking commonalities arise between networks that come from seemingly unrelated fields of science and engineering. This was the birth of complex network science. For decades, specific examples of complex networks have been studied intensely within their unique disciplines. Only recently has there been a surge of interest in studying complex networks in their own right. By bridging the gap between diverse disciplines much progress has been made.

Considerable research has focused on understanding the topology of complex networks and the system properties that give rise to such topologies. The focus of our study, however, is on the dynamics of processes which occur on these complex networks. Although the topology of the network clearly affects

© Jenkinson & Goutsias, 2011
these processes, we will be studying network topology in just enough depth to understand these effects. However, it is important to note that we will only scratch the surface of this highly researched aspect of complex networks. The study of dynamics on complex networks is in its infancy and many questions remain unanswered. We hope that this presentation can serve to bring scientists from diverse backgrounds up to speed on the state-of-the-art techniques in the field complex network science.

1.1 Nonlinear complexity

Before discussing the complications introduced by the topology of complex networks, we preview difficulties introduced by nonlinear components.

One reason why it is not an easy task to study the dynamics of nonlinear complex networked systems is due to the fact that many of these systems have nonlinear processing units – the output of each processing unit is not necessarily a weighted sum of its inputs. Systems comprised of linear components are well understood, even when they are used to build a complex networked system.

For an insight into this issue, let us consider a common linear process: the Fibonacci sequence \( \{0, 1, 1, 2, 3, 5, 8, 13, 21, \ldots \} \). This sequence is built by summing the current value with the previous value to obtain the next value, i.e.

\[
F_{n+1} = F_n + F_{n-1},
\]

is the linear recursion where \( F_n \) is the \( n \)th Fibonacci number. This recursion seems simple enough, but it is of little use if we want to calculate the ten trillionth Fibonacci number. Fortunately, we have the field of linear algebra to deal with such systems. First, we may note that \( F_n \) is not a proper “state variable” since knowledge of \( F_n \) is insufficient to produce \( F_{n+1} \), so we introduce a two-dimensional state variable \( y_n = [F_n \ F_{n-1}]^T \), whose one-step recursion is

\[
y_{n+1} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} y_n.
\]
This implies that
\[
\mathbf{y}_n = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}^n \begin{bmatrix} 1 \\ 0 \end{bmatrix}
= \begin{bmatrix} \frac{1+\sqrt{5}}{2} & \frac{1-\sqrt{5}}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} \left(\frac{1+\sqrt{5}}{2}\right)^n & 0 \\ 0 & \left(\frac{1-\sqrt{5}}{2}\right)^n \end{bmatrix} \begin{bmatrix} \frac{1+\sqrt{5}}{2} & \frac{1-\sqrt{5}}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 0 \end{bmatrix},
\]
which yields the closed form solution
\[
F_n = \frac{1}{\sqrt{5}} \left(\frac{1+\sqrt{5}}{2}\right)^n - \frac{1}{\sqrt{5}} \left(\frac{1-\sqrt{5}}{2}\right)^n.
\]

The point of the previous demonstration is that closed form solutions are available for any linear difference equation. Likewise, for any system modeled by a linear differential equation,
\[
\frac{d\mathbf{y}(t)}{dt} = A\mathbf{y}
\]
we have the closed form solution
\[
\mathbf{y}(t) = \exp(At)\mathbf{y}(0),
\]
where \(\exp(\cdot)\) is the matrix exponential. Similarly, the entire branch of electrical engineering devoted to signal processing begins with methods for dealing with linear time-invariant (LTI) systems. Signal processing uses Fourier analysis which exploits the fact that complex exponentials \(e^{j\omega t}\) are the eigenfunctions of LTI systems. Compared to non-linear systems, linear systems are much easier to deal with analytically and computationally.

Physicists have long known that even small systems can bring about infinitely complex (i.e., chaotic) behavior when nonlinearities are involved. A classical example is the Three Body Problem, in which the motions of three planetary bodies are influenced by mutual gravitational attractions (think of the sun, the earth and the moon). Although this is an example of simple components giving rise to complex behavior, it is not a complex network because the interactions between the bodies are entirely uniform and symmetric (i.e., every “body” interacts with every other “body” according to the same rules). Even the \(n\)-Body Problem, which is the same problem.
but with $n$ planetary bodies, cannot meet our requirements as a complex networked system. This gives some insight into the difficulties involved with nonlinear dynamics on complex networked systems. Before we can examine the dynamics in detail, we need to examine the topological characteristics of complex networks. We begin the discussion by providing a few example systems that will be utilized throughout the class.

## 2 Examples of Complex Networks

### 2.1 Chemical Reaction Networks

Chemical reaction networks are one of the most prolific and important complex networks under investigation. Furthermore, the tremendous generality and flexibility of chemical networks means that most other complex networks can be viewed as a special case of a chemical reaction network when it comes to mathematical and computational analysis. As such, these networks will be our archetypal system of study.

In chemical networks, the components of our system are molecular species. Groups of molecular species, known as reactants, can interact through series of events known as chemical reactions to create new molecular species, known as products. A chemical reaction is called reversible if the reaction can occur in both directions, taking either reactants to products (in the forward reaction) or products to reactants (in the reverse reaction). Note that, in the case of a reversible reaction, the labeling of forward and reverse reactions is arbitrary, but only the forward reaction occurs in an irreversible reaction.

A general chemical reaction system is comprised of $N$ molecular species $X_1, X_2, \ldots, X_N$ that interact through $M$ coupled reactions, given by

$$
\sum_{n \in N} \nu_{nm} X_n \xrightleftharpoons[k_{2m}^{-1}]{k_{2m}} \sum_{n \in N} \nu_{nm}' X_n, \quad m \in M,
$$

where $N := \{1, 2, \ldots, N\}$ and $M := \{1, 2, \ldots, M\}$. Parameters $k_{2m} > 0$, $k_{2m} \geq 0$ are respectively the rate constants of the forward and reverse reactions. These constants tell us how quickly a reaction proceeds. The quantities $\nu_{nm} \geq 0$ and $\nu_{nm}' \geq 0$ are the stoichiometries of the reactants and products. The stoichiometries are whole numbers that tell us how many of each species are consumed or produced in a given reaction. To summarize the topological effect of a chemical reaction network, we often collect the stoichiometries
into an $N \times M$ stoichiometry matrix $\mathbf{S}$ with components $s_{nm} := \nu'_{nm} - \nu_{nm}$, $n \in \mathcal{N}$, $m \in \mathcal{M}$. We call $s_{nm}$ the net stoichiometric coefficient of the $n^{th}$ molecular species associated with the $m^{th}$ reaction.

2.2 Epidemiological Networks

Epidemiological networks study the spread of infectious diseases or agents through a population. The simplest and most widely used model for these networks is known as the SIR model. In this model, an individual in the population can be in one of three states with respect to a disease: susceptible (S), infected (I), or resistant (R). Under this model, there are two types of interactions which an individual may undergo. First, if a susceptible individual comes into contact with an infected individual, the susceptible person can be infected. Second, an infected individual may become resistant if their immune system fights off the infection and confers resistance, or if the individual is killed by the infection. Note that this can be viewed as a special case of a chemical reaction network with $N = 3$ species (S, I, and R) and $M = 2$ reactions:

$$S + I \rightarrow 2I \quad (5)$$

$$I \rightarrow R. \quad (6)$$

If the model is of a slower progressing illness, it may be necessary to include “vital dynamics” (births and deaths) as well. The complexity of a model can be further increased by assuming another layer of networking based on the geographic distributions of populations (since I cannot be infected by someone who is currently in China) and the travel/mixing rates of these populations (if I took an airplane to China to see the Great Wall, I could be exposed by someone who is currently living in China). These so called compartmental models can also include non-geographical “compartments,” such as age brackets of populations who may be affected differently by diseases. In the limit of fine compartmentalization, we can arrive at models that monitor differences on a continuous scale.

Further complexity can be incorporated into the number of states an individual may be in (in our previous example S, I, or R) to more accurately describe the progression of an infection. For example, a person with the flu can be infected but incapable of infecting others since they are getting better as their immune system fights the virus. In this case, we would want to add
the new state to our model (e.g., to describe sick individuals who are not contagious) and include the required interactions that determine how people move in and out of this new state. Additional paths between states may also be possible. For example, if a vaccine is available, an interaction taking individuals from a susceptible state, $S$, to the resistant state, $R$, should be included in Equations 5-6.

An example of a modeling framework that changes the states to fit the disease (i.e., no longer the SIR model) and models more accurately the networks of infectious contacts, is the class of models of sexually transmitted diseases. For many of these infections (e.g., a bacterial infection like syphilis) immunity is not conferred after the patient takes an antibiotic. As a consequence, Equation 6 is replaced by:

$$I \rightarrow S$$

and the only states a person can be in are susceptible and infected. Thus, this model is commonly referred to as the SIS model. Furthermore, if sexual contact is the only way to spread this infection, then the model must take into account the network of sexual partnerships if it is to successfully analyze the spread of the disease.

2.3 Trophic Networks

Trophic networks aim to study the flow of biomass across trophic levels. The goal is to understand how mass and energy are transferred from primary producers (or autotrophs) who generate their own energy from the sun’s rays up to the apex predators who gather their energy and body mass through the consumption of prey lower in the food chain.

Again, this can be viewed as a set of chemical reactions. For an example system, let us assume that we are dealing only with grass ($G$), rabbits ($R$) and wolves ($W$). If we assume the grass supply is approximately constant, since the rabbits fertilize the grass after they eat it, and that rabbits eat grass and produce more rabbits, and that wolves eat rabbits to produce more wolves, and that wolves die at a constant rate, then we have the following system:

$$G + R \rightarrow G + (1 + a_1)R$$
$$W + R \rightarrow (1 + a_2)W$$
$$W \rightarrow \emptyset$$
where $0 < a_1, a_2 < 1$ are constants that detail the percent conversion of biomass and $\emptyset$ is the empty set. Under appropriate modeling assumptions, this system leads to the classical Lotka-Volterra equations, which are differential equations that everyone encounters in a course on dynamical systems (in a math department) or chaos (in a physics department). Again, more complicated models can include geographical considerations and much more complex food webs. However, even in the simplest trophic systems, we can see complex behavior.

Interesting topological details can be gleaned from the nature of the individual interactions. For example, the energetic hypothesis notes that the conversion of biomass is inefficient (e.g. in the above example $a_1$ and $a_2$ are strictly less than 1 and are usually less than 0.1). Because of this any food chain$^1$ is expected to have only a limited number of links (usually 4-5). This is an example of how local characteristics can determine global structure in complex networks.

One tool that we will discuss in detail later is sensitivity analysis, which can determine which components of a system are most crucial to some global behavior. In ecology, there are keystone species that have a disproportionate effect on the structure of the entire ecosystem. Grizzly bears are a classic example, since they transfer biomass from oceanic ecosystems to forest ecosystems – the removal of these bears from the system would be like blowing up a key bridge between two islands. Extinction of a keystone species can produce “cascading extinctions” that result in other species dying out due to the crucial role of the keystone species in the ecosystem. While keystone species have been well documented in ecology, sensitivity analysis has been used in many systems, and it appears that a small number of disproportionately influential components is a nearly universal concept in complex networked systems. For example, the power grid can experience cascading power failures due to a single downed power line, or cancer can develop due to a single mutant protein.

### 2.4 Neural Networks

No discussion on complex networks can be complete without mentioning neural networks in the brain. With 15–33 billion neurons, each having up

$^1$A food chain is an individual path up through the trophic levels of the food web. Note that a single species can be located at multiple trophic levels (e.g. humans eat autotrophs and herbivores).
to 10,000 synaptic connections, there are few complex networks that can compete with the human brain in size and complexity. There are hardly any open problems more intriguing than understanding the human brain, and as such, there is an enormous body of literature surrounding the modeling of biological neural networks. From molecular dynamics of voltage-gated ion channels, to the Nobel Prize winning work by Hodgkin and Huxley on biophysical models of action potential initiation and propagation in neurons, to rapid large-scale mathematical models with no biophysical basis, such as the Izhikevich model, there is no consensus on which level of abstraction will yield the greatest understanding of the system at hand (for a good review, see [1]).

For pedagogical reasons, we choose to focus on a mid-level model that is intuitive enough for novices in neurobiology to understand, and yet rich enough to be a viable candidate for understanding this preeminent complex network. The model consists of \( L \) neurons, with each neuron being in either a quiescent or an active state. Let \( X_{2l-1} \) and \( X_{2l} \) denote a quiescent or active neuron \( l \), respectively. We can assign the following two reactions to the \( l \)th neuron in the network:

\[
X_{2l-1} + \sum_{l' \neq l} \nu_{ll'} X_{2l'} \rightarrow X_{2l} + \sum_{l' \neq l} \nu_{ll'} X_{2l'}
\]

\[
X_{2l} \rightarrow X_{2l-1},
\]

where \( \nu_{ij} \) measures the synaptic weight between neurons \( i \) and \( j \). Note that the first reaction models transition of the \( l \)th neuron from the quiescent to the active state, whereas, the second reaction models transition of the neuron from the active to the quiescent state. As a consequence, we obtain a reaction network with \( N = 2L \) species and \( M = 2L \) reactions.

2.5 Pharmacokinetics

Physiological pharmacokinetic models are used extensively for studying the absorption, distribution, metabolism, and elimination of chemicals and drugs by the body of animals and humans. As a consequence, they are of crucial importance for drug dosing in clinical pharmacology. A large class of pharmacokinetic models is based on the notion of compartmentalization. These models assume the existence of a central compartment (e.g., heart, lungs, brain, etc.), which serves as a site for drug administration to peripheral compartments (e.g., fat, muscles, central nervous system, and liver).
As a specific example, we consider here a previously developed model [6] for studying the effect of tetrachloroethylene, a widely used solvent, on carcinogenesis. This model assumes a division of the human body into the lungs, which serve as the central compartment, and four peripheral compartments, namely fat tissue, poorly perfused tissue (muscles and skin), richly perfused tissue (central nervous system and viscera, except liver), and liver. To model this system, we will denote by \( X_n \) the solvent present in the \( n^{th} \) compartment. Then, we can represent the system by \( N = 5 \) species interacting by the following \( M = 10 \) reactions:

\[
\emptyset \rightarrow X_1 \\
X_1 \rightleftharpoons X_2 \\
X_1 \rightleftharpoons X_3 \\
X_1 \rightleftharpoons X_4 \\
X_1 \rightleftharpoons X_5 \\
X_5 \rightarrow \emptyset .
\]

The underlying reactions model the injection of solvent into lung blood (reaction 1), the exchange of one molecule of solvent between the lung blood and fat tissue (reactions 2 & 3), poorly perfused tissue (reactions 4 & 5), richly perfused tissue (reactions 6 & 7), and liver tissue (reactions 8 & 9), as well as the metabolic clearance of the solvent by the liver (reaction 10). This system is represented graphically in Figure 1.

2.6 Social networks

Recently, interest has emerged in developing mathematical models for social networks that can result in a better understanding of human behavior. In particular, much effort has been devoted to studying the dynamics of social networks, a problem that has been investigated by the physics community many decades ago. Several models for dealing with dynamic processes on social networks are currently available, with many of them fitting nicely into the current framework. As an example, we focus on opinion spreading in social networks, a process of political, marketing, and general sociological interest, by utilizing concepts from a recently proposed heterogeneous voter model [7].

Let us consider a population of \( L \) individuals whose state (opinion) takes binary 0/1 values. Possible opinions abound: democrat or republican, for
Figure 1: A graphical representation of the pharmacokinetic model for tetra-
chloroethylene. Black circles are reactions and white circles represent species. 
See Section 3.3 for details on this type of graph.

or against an idea, like or dislike of a product, etc. Let $X_{2l-1}$ and $X_{2l}$
denote the $l^{th}$ individual with opinion 0 and 1, respectively. We can model
interactions between the $N = 2L$ individuals by employing the following
$M = 2L$ reactions:

$$
X_{2l-1} + \sum_{l' \neq l} \nu_{ll'} X_{2l'} \rightarrow X_{2l} + \sum_{l' \neq l} \nu_{ll'} X_{2l'}
$$

$$
X_{2l} + \sum_{l' \neq l} \nu_{ll'} X_{2l' - 1} \rightarrow X_{2l-1} + \sum_{l' \neq l} \nu_{ll'} X_{2l' - 1}, \quad \text{for } l = 1, 2, \ldots, L,
$$

with each pair corresponding to the $l^{th}$ individual, where $\nu_{ij}$ is a number
between 0 and 1 that measures the weight of social influence of the $i^{th}$ indi-
vidual on the $j^{th}$ individual. Clearly, the $2l - 1$ reaction models change in the
$l^{th}$ individual’s opinion from 0 to 1, whereas, the $2l$ reaction models opinion
change of the same individual from 1 to 0.
2.7 Multi-agent systems

The study of multi-agent networks focuses on systems in which many intelligent agents, such as autonomous vehicles that observe and act upon their environment, interact with each other to achieve a certain goal. To illustrate the fact that multi-agent systems can also be modeled as a reaction network, we consider here a system comprised of \( L \) autonomous unmanned vehicles (AUV’s) that can move over a two-dimensional bounded rectangular space in a discrete fashion [5]. For simplicity, we assume that, at each step, an AUV located at a discrete point \((i, j)\) in space can move towards one of four possible directions, namely east to point \((i + 1, j)\), west to point \((i - 1, j)\), north to point \((i, j + 1)\), or south to point \((i, j - 1)\). We want to develop a mathematical approach that can be used to describe vehicular motion so that the AUV’s reach a spatial configuration \(x\) at steady-state with desired probability \(q(x)\). In particular, \(q(x)\) could assign high probability over configurations that maximize a given design objective and low or zero probability over the remaining configurations.

In the following, we employ two species \( X_{2l-1} \) and \( X_{2l} \) whose populations \( x_{2l-1} \) and \( x_{2l} \) denote the position of the \( l \)th AUV on the two-dimensional rectangular grid. For example, if the \( l \)th vehicle is located at point \((i, j)\) on the grid, then \( x_{2l-1} = i \) and \( x_{2l} = j \). We can now characterize the motion of all AUV’s in the multi-agent network under consideration by \( N = 2L \) species interacting through the following \( M = 4L \) reactions:

\[
egin{align*}
X_{2l-1} + X_{2l} + \sum_{l' \neq l} (X_{2l'-1} + X_{2l'}) & \rightleftharpoons 2X_{2l-1} + X_{2l} + \sum_{l' \neq l} (X_{2l'-1} + X_{2l'}) \\
X_{2l-1} + X_{2l} + \sum_{l' \neq l} (X_{2l'-1} + X_{2l'}) & \rightleftharpoons X_{2l-1} + 2X_{2l} + \sum_{l' \neq l} (X_{2l'-1} + X_{2l'}) .
\end{align*}
\]

The first two reactions model one-step motion of the \( l \)th AUV towards east/west, whereas, the other two reactions model one-step motion towards north/south.

3 Graph-Theoretic Representations

3.1 Directed and Undirected Graphs

We now present a high level “crash course” on graph theory that will facilitate our discussion on the topological structure of complex networks.
A graph, \( G \), is a mathematical construct composed of two sets \((V, E)\). \( V \) is a countable set whose elements are referred to as nodes or vertices or (in Petri nets, which we will discuss shortly) places. For simplicity, we can always label the nodes using the positive natural numbers, so that \( V := \{1, 2, \ldots, |V|\} \), where \(|V|\) is the cardinality of the set \( V \) (i.e., the number of its elements). \( E \) is a set of edges, which are pairs of nodes from the set \( V \). If these pairs are unordered sets, then the graph \( G \) is called undirected. When the edges are ordered pairs, or 2-tuples, the graph \( G \) is called a directed graph or a digraph.

Nodes that are connected by edges in \( E \) are referred to as nearest neighbors or adjacent nodes. All information about the graph \( G \) is conveniently contained in the \(|V| \times |V|\) adjacency matrix \( A \), with elements \( \{ a_{ij}, i, j \in V \} \), given by

\[
a_{ij} = \begin{cases} 
1, & \text{if } (i, j) \in E \\
0, & \text{if } (i, j) \notin E,
\end{cases}
\]

so that \( A \) is akin to an indicator function denoting the support of \( E \). Note that, for an undirected graph, \( A = A^T \), but this is not true for a digraph. Visual representations of graphs commonly represent nodes with labeled circles and edges as lines or arcs connecting appropriate nodes. In directed graphs, the edges have arrow heads to specify the directionality of the connection. Two example graphs and their corresponding adjacency matrices are depicted in Figure 2.

A path, \( p_{i_1,\ldots,i_{n+1}} \), is an ordered sequence of nodes \((i_1, \ldots, i_{n+1})\), such that \( i_\rho \in V \), for \( \rho = 1, \ldots, n + 1 \), and \((i_\epsilon, i_{\epsilon+1}) \in E \), for \( \epsilon = 1, \ldots, n \). The length of a path is given by the number of edges traversed in order to go from the initial node to the final node of the path (i.e., \( n \) is the length of the path \( p_{i_1,\ldots,i_{n+1}} \)). If a path exists between two nodes, then we say that these nodes are connected. For example, a path of length \( n = 3 \) in the undirected graph depicted in Figure 2 would be \((2, 4, 3, 1)\), but this is not a valid path in the directed graph. Let \( N_{ij}(n) \) denote the number of paths of length \( n \) connecting the \( i^{th} \) node to the \( j^{th} \) node. We may calculate this quantity using powers of the adjacency matrix, since

\[
N_{ij}(n) = [A^n]_{ij},
\]

where \([ \cdot ]_{ij}\) takes the \((i, j)^{th}\) component of a matrix and \( A^n \) is the \( A \) matrix multiplied by itself \( n \) times. The shortest path length between the \( i^{th} \) and \( j^{th} \) nodes is denoted by \( \ell_{ij} \). If no path exists between the \( i^{th} \) and \( j^{th} \) nodes, then we set \( \ell_{ij} = \infty \).
The degree of the $i$th node in a graph is defined to be the number of edges, $k_i$, that connect to that node. This definition is unambiguous in undirected graphs. However, in directed graphs, it is common to assign two numbers: the in-degree, $k_{in,i}$, and the out-degree, $k_{out,i}$, which are the number of edges entering the $i$th node and the number of edges leaving the $i$th node, respectively. Generally, the sum of these two numbers denotes the degree of a node in a directed graph (i.e., $k_i = k_{in,i} + k_{out,i}$). A common tool used to characterize graphs is the degree distribution, $p(k)$, which tells us the probability that a randomly selected node will have degree $k$. Directed graphs may also have an in-degree distribution as well as an out-degree distribution. Statistical characterizations of a graph are possible using degree distributions as we may take expectations (e.g., to find the average degree of a graph) or calculate higher-order moments. Note that, for a directed graph, we find identities such as $E[k_{in}] = E[k_{out}] = \frac{1}{2}E[k]$, where $E[\cdot]$ denotes expectation.
since every edge leaving one node must enter another node.

There are many other quantities that can be used to describe the topology of a graph. If you are interested, you may want to investigate terms such as degree centrality, betweenness distribution, clustering coefficient, cyclomatic number, cliques, subgraphs, modules, graph diameter, vertex eccentricity, and Pearson assortativity coefficient (e.g., see [3]). We will not delve deeper into such issues, since our focus is on the processes occurring on graphs, and there are numerous sources available on graph theory and on the topology of complex networks.

3.2 The “Population” Simplification

Suppose we want to make a graph that represents the mating possibilities for five caged lions in a zoo (who have been given human names to help keep their sexes straight): Matt, John, Steve, Katie, and Mary. Figure 3 is the resulting graph if each lion is assigned to a node on the graph, with edges connecting possible mating pairs. Note that there is a symmetry present. Namely, Katie and Mary are connected in identical ways, and then Matt, John, and Steve also have identical connections. Thus, we can collapse the graph to have just two nodes (representing these two classes of individuals) and one edge (representing the edges that originally went from every member of one group to every member of the other group). In general, to retain the information stored in the original graph, we will assign a state variable to each class node that enumerates how many individuals (i.e., the population) from the original graph are being stored in this new node.

This simplification is crucial in the study of complex networks, especially when the number of individual components in a system is very large. Think of the simple case of pouring vinegar (with acetic acid) into a science fair volcano full of baking soda (sodium bicarbonate). There are on the order of $10^{23}$ molecules of each chemical and therefore approximately $10^{46}$ edges in this graph. Although each molecule of sodium bicarbonate may be different, it behaves in the exact same way with any other molecule of acetic acid. Thus, we can simplify this extraordinarily complex situation into a much more manageable task of having two nodes, one edge, and the responsibility of polling how many individual nodes have been collapsed into these two new class nodes (i.e., by measuring the concentration of our two chemicals).
Figure 3: An example of “population” simplification which combines nodes that behave identically within a system. In this case, we exchange graphical simplicity for added complexity of the state space to store how many individual nodes are collapsed into each class node (in this case, three nodes are collapsed into the male class and two nodes are collapsed into the female class).

3.3 Weighted Graphs and Bipartite Graphs

The mathematical abstraction of a graph is often useful when describing real world phenomena. Unfortunately, if we were to describe complex networks as a simple collection of edges and nodes, much important information would be lost in the process. For example, Facebook will gladly make a graph for you with the nodes being your “friends” and the edges connecting individual “friends” who are in contact with each other on facebook. Unfortunately, all edges in this network are homogeneous – they represent only one type of connection, namely the existence of contact between individuals. In reality,
however, the amount of contact between individuals may vary significantly. One individual may be your best friend, or perhaps even your significant other, in which case contact may be very frequent, while another may be someone you met once at a party. A natural extension to our graph-theoretic description that can account for such heterogeneity is to assign weights to each edge in our network. Thus, your best friend or significant other would have a larger weight in your social network than your other acquaintances.

This extension is rather simple to incorporate into our framework. An edge is now associated with a weight (which can be real-valued) detailing the strength of connection between two nodes in the network. The $|V| \times |V|$ binary valued adjacency matrix $A$ may be now replaced by a $|V| \times |V|$ real valued weighting matrix, $W$, with its $(i,j)$th component being the weight $w_{ij}$ of the edge connecting the $i$th node to the $j$th node (i.e., $[W]_{ij} = w_{ij}$ and $w_{ij} = 0$ if there is no edge from node $i$ to node $j$). We have $W = W^T$ for undirected graphs only. Furthermore, a natural counterpart of the degree of a graph is the node strength, $s_i$. For undirected networks, we have that $s_i = \sum_{j \in V} w_{ij}$. For directed networks, we can define the in-strength $s_{in;i} = \sum_{j \in V} w_{ji}$, the out-strength $s_{out;i} = \sum_{j \in V} w_{ij}$, and the strength $s_i = s_{in;i} + s_{out;i}$. Depending on the interpretation of a negative weight within a specific context, one may wish to alternatively define strength using the absolute value of the weights.

Even a weighted graph, however, is not sufficiently descriptive to epitomize a general complex networked system. Take, for example, a typical catalytic reaction scheme, with $X$ and $Y$ being reactants, $W$ being a catalyst and $Z$ being a product,

\begin{align*}
X + W & \rightarrow XW \quad (15) \\
Y + XW & \rightarrow XYW \quad (16) \\
XYW & \rightarrow WZ \quad (17) \\
WZ & \rightarrow W + Z. \\
\end{align*}

As an exercise, attempt to capture Equations 15-18 using the graph theoretic principles presented thus far. There are many options on how to do this, and indeed an entire textbook exists on the subject [4]. Unfortunately, the tools presented thus far (which are in fact the tools most generally used in the literature) are incapable of capturing all information present in Equations 15-18. Therefore we must extend our toolbox one step further in order to fully describe general complex networked systems.
Let us first conceptualize a further generalization that will help us describe such complicated systems, and then we can proceed to formalize this concept mathematically. At the heart of complex networks we have two fundamental entities: *interacting components* (e.g., chemical compounds or animal species) and *processing units*, which tell us how these components interact with each other (e.g., chemical reactions or predator-prey relationships). We have attempted to use graph theory to assign nodes to interacting components and edges to the processing units. But how can we represent Equation 15 in this way? If \( X \), \( W \), and \( XW \) are our three interacting components, then each is assigned to a node. The chemical reaction in Equation 15 is the processing unit, which tells us to use edges to connect these nodes. We are left with a fully connected graph with three nodes and no understanding that \( X \) and \( W \) combine to create \( XW \). This is because, even with weights, the edges of our graph are homogeneous. What if we give processing units their own nodes, instead of simply representing them as edges? However, we would like to distinguish processing units from interacting components, so we could assign processing units a different kind of a node than the interacting components. Now, it is simple to describe Equations 15-18, as can be seen in Figure 4.

What we have described conceptually is a well defined mathematical object known as a *bipartite graph* or a *bigraph*. A bigraph is the triplet \( B = (C, P, E) \), where \( C \) is a countable set of nodes as before, \( P \) is a new and different countable set of nodes, and \( E \) is a set of edges that may only connect nodes in \( C \) to nodes in \( P \), and vice versa (i.e., an edge cannot connect a node in \( C \) to another node in \( C \), and likewise for \( P \)). The edges may be directed or undirected, and weighted or weight-free as before. The sets \( C \) and \( P \) are used to represent interacting components and processing units, respectively.

### 3.4 Petri nets

We now discuss a high level framework that is capable of encompassing most (if not all) complex networked systems, and naturally leads to various methods for modeling the dynamics on these networks.

A *Petri net* is a quadruple \( P = (C, P, E, x_0) \), where \( (C, P, E) \) defines a weighted directed bipartite graph and \( x_0 \) is known as the initial marking of the graph [9]. Since the sets \( C \) and \( P \) are finite, we can define the number of places, \( N = |C| \), and the number of transitions, \( M = |P| \). Moreover, we can appropriately relabel the sets \( C \) and \( P \) to correspond to the simpler natural
numbered sets \( \mathcal{N} := \{1, 2, \ldots, N\} \) and \( \mathcal{M} = \{1, 2, \ldots, M\} \). All information about \( E \) can be stored in two \( N \times M \) stoichiometry matrices, \( V \) and \( V' \), with components, \( \nu_{nm} \) and \( \nu'_{nm} \), respectively. The component \( \nu_{nm} \) is the weight of the edge leaving the \( n \)th place to enter the \( m \)th transition (it is zero if there is no edge connecting the two), and likewise, \( \nu'_{nm} \) is the weight of the edge entering the \( n \)th place from the \( m \)th transition. The initial marking \( x_0 \) is an \( N \times 1 \) vector that assigns an initial state to each place in \( C \). The state \( x(t) \) [which satisfies \( x(0) := x_0 \)] defines the number of tokens at each place in the graph at time \( t \).

Modeling the dynamics of a complex network is now equivalent to modeling the flow of tokens on the Petri net [i.e., finding \( x(t) \), for \( t \geq 0 \)]. In
particular, each transition (or processing unit) in $P$ is assigned a firing rule that defines when a transition will occur. Note that the firing rule may be stochastic. When the $m$th processing unit fires, it instantly takes $\nu_{nm}$ tokens from the $n$th place (for all $n \in N$) and distributes $\nu'_{nm}$ tokens to the $n$th place (again for all $n \in N$). Thus, it is common to define the stoichiometry matrix $S$, with components $s_{nm} := \nu'_{nm} - \nu_{nm}$, which tell us the net change of the state $x_n$ of the $n$th place due to the $m$th transition firing once. Thus, if the $m$th transition fires once during $[t, t + dt)$, and it is the only transition that fires during that time, then

$$x(t + dt) = x(t) + s_m,$$

(19)

where $s_m$ is the $m$th column of $S$.

The following table lists example complex networks and possible Petri net representations of such networks.

<table>
<thead>
<tr>
<th>Network</th>
<th>Places (Interacting Components)</th>
<th>Transitions (Processing Units)</th>
</tr>
</thead>
<tbody>
<tr>
<td>biochemical</td>
<td>molecules</td>
<td>chemical reactions</td>
</tr>
<tr>
<td>biological neural</td>
<td>neurons</td>
<td>synaptic transmissions</td>
</tr>
<tr>
<td>epidemiological</td>
<td>healthy, infected individuals</td>
<td>infectious contacts</td>
</tr>
<tr>
<td>world wide web</td>
<td>web pages</td>
<td>hyper-links</td>
</tr>
<tr>
<td>trophic</td>
<td>predators, prey</td>
<td>consumption, predation</td>
</tr>
<tr>
<td>power grid</td>
<td>generators, end users</td>
<td>power lines, substations,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>transformers</td>
</tr>
<tr>
<td>collaborative</td>
<td>professionals</td>
<td>collaborations</td>
</tr>
<tr>
<td>social</td>
<td>people</td>
<td>contacts, relationships</td>
</tr>
<tr>
<td>internet</td>
<td>computers</td>
<td>routers, ethernet cables</td>
</tr>
</tbody>
</table>

### 3.4.1 Chemical Petri Nets

As an example of a Petri net, let us consider the chemical network given by Equations 15-18. For this network, we have $C = \{X, Y, W, Z, XW, XYW, WZ\}$ and $P = \{Rxn1, Rxn2, Rxn3, Rxn4\}$, with Rxn2 and Rxn3 being reversible transitions (i.e., transitions that can occur in the forward or backward direction, which update the states by $s_m$ or $-s_m$, respectively). Note that Figure 4 depicts our weighted directed bipartite graph $B = (C, P, E)$, where by convention weights equal to unity are not displayed on the corresponding
edges. By relabeling $C$ using $N$, (i.e., by setting $X_1 = X, X_2 = Y, X_3 = Z, \ldots, X_7 = WZ$) we can define the following stoichiometry matrices:

$$
V = \begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 \\
0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0
\end{bmatrix},
V' = \begin{bmatrix}
0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 1 & 1 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{bmatrix},
$$

and

$$
S = \begin{bmatrix}
-1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
-1 & 0 & 0 & -1 \\
0 & 0 & 0 & -1 \\
0 & 1 & 1 & 0 \\
0 & -1 & 1 & 0 \\
0 & 0 & -1 & 1
\end{bmatrix}.
$$

Matrices $V$ and $V'$ capture all topological information of the Petri net (i.e., they can be considered as defining $E$). If we assign initial molecular numbers (i.e., the number of each molecule present at time $t = 0$) to all seven species in $C$, and stored them in a vector $x_0$, then we have our Petri net fully specified as $P = (C, P, E, x_0)$.

If we define a rule detailing when each reaction occurs (i.e., a firing rule for each processing unit in $P$), then we can model the system by updating the state vector $x(t)$ at time $t$ according to the $M \times 1$ vector $z(t)$, whose components $z_m(t)$ define the net number of times the $m^{th}$ transition occurs in the forward direction within the time interval $[0, t)$. If the reverse reaction occurs, we count this as a forward reaction occurring $-1$ times. Thus, $z_m(t)$ can be positive or negative when the $m^{th}$ reaction is reversible, and $z_m(t) \geq 0$ when the $m^{th}$ reaction is irreversible. As a consequence, we can calculate the state $x(t)$ at any time $t \geq 0$ if we know $z(t)$ and $x_0$ using the following equation:

$$
x(t) = x_0 + S z(t), \quad t \geq 0.
$$

It is common to refer to $z_m(t)$ as the degree of advancement of the $m^{th}$ processing unit (reaction in the case of a chemical network).

4 Network Topology

Discussions on complex network topology are usually simplified by focusing on standard (i.e., not bipartite) graphical representations of networks. One can “project” any bipartite graph $B = (C, P, E)$ onto a standard graph $G = (C_0, E_0)$ by taking $C_0 = C$ and by including an edge in $E_0$ only when the nodes in $C_0$ connected by this edge were originally nodes in $C$ that were
connected to the same node in $P$. In the new simplified graph $G$, interacting components are represented as nodes in $C_0$ and interactions are represented by the edges in $E_0$.

### 4.1 Scale-free distribution

One of the earliest discoveries about complex networks was that their degree distribution, $p(k)$, is usually characterized by “heavy tails.” This means that real world networks have an unusually high probability of containing nodes with a large number of connections (such nodes often represent hubs). Specifically, it has been empirically found that an astounding variety of real world networks are *scale-free*, meaning that their degree distribution follows a power law of the form:

$$p(k) \sim k^{-\gamma},$$

where $\gamma$ is a constant usually (though not always) taking values between 2 and 3. Examples of scale-free networks include, but certainly not limited to, metabolic, protein, language, social, collaborative, trophic, and transportation networks [3, 8].

Special care must be taken, however, when one claims that a network is actually well modeled by a scale free network. A classic example is the internet, which is popularly believed to be a preeminent scale free network. In fact, this belief is incorrect despite a large volume of work appearing in top journals by top scientists spreading this misconception (see [10] for an extremely interesting discussion of this scientific plague). In this case, researchers neglected to account for the fact that there is a *physical system* being modeled. Routers cannot have an arbitrary number of connections, and the scale-free models predicted routers with far more connections than was physically feasible with the technologies in place.

This is a theme of much of the authors’ work, and will show up in many places in the text. Mathematical models are great, but when we are modeling a physical system, it is crucial that our model obeys the fundamental constraints (e.g. those of thermodynamics) that govern physical systems. Furthermore, a model’s simplifying assumptions and their inherent limitations should be well understood so that erroneous conclusions may be avoided.
4.2 Preferential attachment

While many explanations have been offered as to why the degree distribution of so many diverse networks follows the power law given by Equation (14) (e.g., see pages 9-10 in [11]), perhaps the most discussed explanation is a self-organizing property known as preferential attachment. This property refers to the fact that a newly added node to the system may be more likely to form an edge with a node that already has a large number of attached nodes. This rich-getting-richer type of phenomenon will produce scale-free random graphs under suitable conditions [12].

4.3 It is a small, robust world

The existence of hubs is responsible for many behavioral features of complex networks. For example, the small world property of complex networks can be quantified by the fact that these networks have an average shortest path length between two nodes that scales as \( \log(N) \), where \( N \) is the number of nodes in the graph. Compare this with the fact that a \( d \)-dimensional lattice has an average shortest path length that scales as \( N^{1/d} \). Clearly, the presence of hubs that connect to a large number of nodes (and thus serve as shortcuts) ensure that most nodes maintain a short path between each other. Everyone has probably experienced this phenomenon first-hand. Games like six degrees of Kevin Bacon, or six degrees of Wikipedia (where a person tries to get from one Wikipedia page, such as “teddy bear,” to a completely unrelated page, such as “nuclear weapon,” by clicking less than 6 hyper-links within each article\(^2\)) take advantage of this property. Also, most people have at one point sat next to a complete stranger on a plane only to find some seemingly amazing “small world” mutual connection – such experiences are artifacts of the scale-free distribution in social networks.

Another property of scale-free networks that arises from complex topological features has to do with the robustness/fragility of the network. Scale-free networks are incredibly robust to random node failures and, at the same time, incredibly vulnerable to targeted attacks. How can this be? The answer is

\(^2\)I found a path in 4 clicks by going from “teddy bear” to “united states president” to “united states armed forces” to “world war 2” to “nuclear weapon.” Can you find a shorter path? Note that my three intermediate pages are hubs that contain links to many different subjects, explaining how a shortcut can exist between such disparate subjects as a child’s toy and a weapon.
hubs. Random failures of web-servers around the world leave the World Wide Web almost completely unaffected, even though it happens by the second. On the other hand, if a targeted attack occurs on Google’s servers (most definitely hubs of the World Wide Web), one would expect massive effects to ripple through the network. Sensitivity analysis is one mathematical tool that we will later utilize to understand and exploit these properties.

References


