Chapter 5

Growing Random Networks

Another prominent class of models of random networks consists of those where new nodes are born over time and form attachments to existing nodes when they are born. As an example, consider the creation of a new web page. When the web page is designed, it will often include links to existing web pages. The web page might be updated over time; but nonetheless, a nontrivial portion of its links will have been included when the web page is first created. Over time, an existing page will be linked to by new web pages. The same is true of people entering a school, or a new job or new neighborhood. This means that there is a fundamental difference between a growing network and a static one. Time introduces a natural heterogeneity to nodes based on their age in a growing network. This heterogeneity is important for two reasons. First, it is present in many applications where individuals or nodes enter (and leave) networks over time. This includes the web page example mentioned above, as well as many networks of friendships, acquaintances, citations, and professional relationships. People enter and leave networks and accumulate connections over time. Second, the added heterogeneity comes in a simple form that allows the model to move beyond the Poisson random networks of Erdös and Rényi and to have extra layers of richness; but still be manageable to analyze.

In this chapter, I discuss growing random network models and show some of the properties that emerge. The first aspect that will be important is that the resulting degree distributions will be richer than that of Poisson random networks, and at one extreme will provide an explanation for how scale-free distributions might naturally emerge. This contrasts with models such as the configuration model, where we essentially program in the degree distribution that we want and then generate a network. The configuration model is a workhorse for studying diffusion and other properties on

a random network, but precisely because it is simple and lacks the full richness of more foundational or grounded models of networks. Growing networks will allow for several key factors that generate network characteristics that match some of the observations about social networks discussed in Chapter 3. First, when nodes are born we can consider different ways in which they attach to existing nodes. At one extreme, where newborn nodes pick nodes to link to uniformly at random, we will just have a growing variation on an Erdös-Rényi random network. At the other extreme, where they pick nodes in proportion to the current degrees of the existing nodes (something named "preferential attachment" by Barabasi and Albert [40]), nodes that are older and have had a chance to grow in degree will grow faster than younger nodes who have lower degrees. This sort of rich-get-richer process leads to scale-free distributions. Preferential attachment has a nice interpretation in that if we randomly pick a node and then start searching through the network, we will end up finding nodes in proportion to their degree. We can also consider hybrid models, where the attachment probabilities vary between these extremes. Despite claims in the literature that many social networks exihibit power-laws (having scale-free degree distributions), by fitting such hybrid models it becomes clear that most social networks lie between these extremes.

Beyond degree distributions, growing random networks provide insight into other observed characteristics. By the nature of the process, older nodes will have higher degrees on average, and, since older nodes have a greater proportion of their connections to older nodes than do younger nodes, a natural positive correlation in degrees emerge. Actually, not only do we see correlation in degrees, but also age-based homophily which is also consistent with many observed social networks. To the extent that the process involves some form of preferential attachment, we also see large hub nodes emerge in the networks, and this will produce a lower diameter than in a Poisson random network. Finally, certain variations of hybrid models produce the high clustering observed in many social networks. This comes out in hybrid models where newborn nodes find the nodes that they link to by navigating the network itself. That is, to the extent that people are introduced to someone who becomes their friend through a friend, then we see natural clustering emerge. The same is true in citation networks where one finds new articles to cite by examining the bibliography of other relevant articles; and similarly in finding new web pages to link to by following links of pages one is already linked to, and so forth. In short, the growing random networks introduce new aspects to network formation that generate structures exhibiting networks that more closely match many observed characteristics of social networks.

5.1 Uniform Randomness: an Exponential Degree Distribution

To get a feeling for a process where a network grows over time, I start by discussing a dynamic variation on the Poisson random network model, where nodes are born over time and form links with existing nodes at the time of their birth.[1]

As nodes are born over time, index them by the order of their birth. Thus, node i is born at date i, where $i \in \{0, 1, 2 \ldots\}$. I postpone the question of how we should interpret the scale of time until we investigate situations in which nodes are born in clusters. For now, think of each period of time as indicating that a new node has been born, regardless of how much physical time has passed since the last node entered the system.

A node forms links to existing nodes when the new node is born. To start, examine the case where links are undirected. Let $d_i(t)$ be the degree of node i (born at time i) at a time t. So, $d_i(i)$ is the number of links formed at a node's birth, and then $d_i(t) - d_i(i)$ will be the number of links that node i gets from the new nodes that were born between time i and time t.

Consider a variation of the Poisson random setting, where each newborn node randomly selects m of the existing nodes and links to them. To have things well defined, start the network with $m+1$ nodes born at times $\{0, 1, \ldots, m\}$, each connected to each other. The specifics of this will not be of great consequence when we look at limiting properties of the system, but it is helpful in order to be able to properly analyze the system. Thus, the first newborn node that we consider is the one born at time $m + 1$.

At the end of time $m + 1$, m of the older nodes will have new links and one older node will not; while the newest node will have m links. Each of the pre-existing nodes expects to gain $m/(m + 1)$ links (or one link with a probability of $m/(m + 1)$). At the end of time $m + 2$, there are different possibilities: m of the $m + 2$ pre-existing nodes will have gained a new link, while 2 of them will not have. Depending on which 2 do not gain a link we have different possibilities for degree distributions that could be realized. As we continue, the number of possible realizations of the degree distribution grows. While it is hard to keep track of the potential realizations and their relative probabilities, we can do some more direct calculations. If we look at time t, a node i

[1] Here, we already see one of the deficiencies of such models, as the only links formed over time are between a new node and an existing node. There are no new links formed between existing nodes over time. We can return to this feature later.

born at time $m \leq i < t$ has an *expected* degree at time t of

$$m + \frac{m}{i+1} + \frac{m}{i+2} + \cdots + \frac{m}{t},$$

or

$$m\left(1 + \frac{1}{i+1} + \frac{1}{i+2} + \cdots + \frac{1}{t}\right). \tag{5.1}$$

For large t, this is approximately[2]

$$m\left(1 + log\left(\frac{t}{i}\right)\right). \tag{5.2}$$

Therefore, although it is difficult to deduce the actual degree distribution of this network formation process, it is relatively straightforward to deduce the distribution of *expected* degrees. For a large t, the nodes that have expected degree less than d are (using the approximation) those such that

$$m\left(1 + log\left(\frac{t}{i}\right)\right) < d. \tag{5.3}$$

We rewrite this as the nodes i such that

$$i > te^{1-\frac{d}{m}}.$$

Thus, the nodes with expected degree less than d (where $d < m\left(1 + log\left(\frac{t}{m}\right)\right)$) are those born after time $te^{1-\frac{d}{m}}$. This is a fraction of $1 - e^{1-\frac{d}{m}}$. Thus we have deduced an approximation of the distribution function of the expected degrees at time t. For $d < m\left(1 + log\left(\frac{t}{m}\right)\right)$, the fraction of nodes with expected degrees less than than d is

$$F_t(d) = 1 - e^{-\frac{d-m}{m}}. \tag{5.4}$$

This is a variation of an exponential distribution. In particular, each node starts out with m links, and then the expected links that a random node expects to gain over time has an exponential distribution with expected value m.

Note that the distribution is in fact independent of time t. This is because the faction of nodes with no more than some degree d is actually the same over time. As more nodes are being born, more of them are also below any given level. This can be seen in that the expected degree of a given node in (5.2) is dependent on t/i, so it depends on when i was born relative to the overall set of nodes, and not on i's absolute date of birth. This is true of some processes and not others.

[2] $\sum_{k=1}^{n} \frac{1}{k} = H_n$ is a Harmonic Number and has various approximations, as discussed in Section **??**. For large n, an approximation of H_n is $\gamma + log(n)$, where γ is the Euler-Mascheroni constant which is roughly .577; and the difference between this approximation and H_n tends to 0. Thus, we can rewrite (5.1) as $m(1 + H^t - H^i)$, which then leads to the stated approximation.

5.1.1 Mean-Field Approximations

Above, we calculated the distribution of *expected degrees* after time t. How close is this to the distribution of *actual degrees* after time t? This turns out to be a good approximation in this particular model. While proving that this is a good approximation of the actual degree distribution is beyond the scope of this text, let me outline the issues. First, the distribution we derived can easily be shown to be an accurate approximation of the distribution of expected degrees for large t. Our only approximations in this respect came in the rounding the sum of the Harmonic Series, and in not worrying about the differences between the first m nodes and other nodes, which is only an issue for a vanishing fraction of nodes and would only enter the calculation if we examined the very high tail of the distribution. Thus, the place where the approximation might face difficulties is in the difference between a distribution of expected degrees and actual degrees. Here we benefit from the fact that the relative distance of the degree of a node from the expected degree (actual degree minus expected degree divided by expected degree) is going to 0 for this process. The degree of a given node can be seen as the realization of a sum of independent random variables (whether or not each newborn node happens to link to it). While these independent random variables have different probabilities and expectations, their summed expectation relative to the expected value of any given one tends to infinity, and so by a variation on the law of large numbers, (e.g., see Landers and Rogge [400]) we can deduce that the *ratio* of the actual degree relative to the expected degree of any given node tends to 1. There is still much work to be done to establish that the distribution over nodes then converges, as one has to aggregate across nodes. If we fix some degree d, and then ask how many nodes end up on the other side of it compared to the nodes degree, this probability vanishes for any given node, since its expected degree only increases and its ratio over the expected degree goes to one. The key to aggregating, is then showing that this convergence has some uniformity in it, so that the fraction of nodes whose ratios of realized to expected degrees are off by more than a given amount is going to 0.

When we compare the network that emerges from this growing system to a Poisson random network there are two differences. First, in this process, each node starts with a given number m of links. Then it is only the additional links that are random. Thus, an appropriate benchmark random network would not be the Poisson random network of Erdös and Rényi, but instead a variation where there are t nodes, and each picks m

others at random to link to.[3] There each node would approximately have a degree of m plus a Poisson random variable with expectation m. The main difference between the distribution for a Poisson random network and exponential distribution from the growing random network is that the exponential distribution has more of a spread to it: The older nodes tend to have higher degrees and the younger nodes have lower degrees.

The previous section gave us a first exposure to what is known as a *mean-field approximation*. The full randomness of the process was quite complex, and so rather than try to deduce the degree distribution of the process directly, we found an approximation of the distribution of expected degrees. We then argued why, at least in that specific case, the distribution of expected degrees is a valid approximation of the actual degree distribution.

Random graph processes tend to be complex; especially ones where heterogeneity enters the system through time so that nodes are facing different distributions on how connected they will be. As such, deriving the degree distribution at any point in time can be quite difficult. A standard technique, borrowed from the statistical-physics literature, for solving such complex dynamic systems is to use a mean field approximation. That is, one assumes that the system evolves so that things occur at the average level rather than randomly. For instance, let us suppose that there are already 100 nodes and a new node appears and is supposed to form links to existing nodes independently with probability 1/10. Under such an approximation, we instead suppose that the node forms exactly 10 links. This, coupled with a continuous time approximation, allows us to model the change in time of a given node's degree at a fixed rather than a stochastic rate. All of the variation that comes into the system under such an approximation is not due to the stochastics, but rather through other forms of heterogeneity in starting conditions - for instance, here the heterogeneity is in terms of dates of birth. While there are many obvious departures from reality in such an approximation, in many situations these techniques provide remarkably accurate estimates.

There is still distressingly little that we know analytically about when such approximations are good and when they are not. The analysis above provides an argument for why this is an accurate approximation for large growing networks with uniform probabilities on attachment; but such arguments become more difficult with more complex

[3] Here we would have to either admit duplicate links between pairs of nodes, or ignore them.

processes and generally are overlooked entirely.[4] A standard (but not fully satisfactory) technique to verify the accuracy of the approximation is to compare the approximations to simulations of the actual process for some range of parameter values. While we might prefer to be able to calculate things like degree distributions directly, this usually turns out to be intractable for all but the starkest of models, and hence turning to mean-field approximations is a next-best alternative.

5.1.2 Continuous Time Approximations of Degree Distributions

Let us now re-examine the growing random network analyzed in Section 5.1, but do so with an alternative technique, working with a continuous time mean-field approximation.

A new node is born at time t. It forms m links by uniformly randomly picking m out of the t existing nodes.

Node i's degree is thus described by a starting condition of $d_i(i) = m$ and an approximate change over time of

$$\frac{dd_i(t)}{dt} = \frac{m}{t},$$

for each $t > i$. This is due to the new node born at each time spreading it's m new links randomly over the t existing nodes at time t.

This differential equation has a solution

$$d_i(t) = m + m \log\left(\frac{t}{i}\right).$$

Now, from this we can derive an approximation of the degree distribution. We again note that the degrees of nodes are increasing over time. So for instance, if we ask how many nodes have degree of no more than 100, and we see that a node born at time τ has degree of exactly 100, then we are equivalently asking how many nodes were born on or after time τ. So, if it is currently time t, then the fraction of nodes having degree of no more than 100 would be $(t - \tau)/t$. In this manner we derive a degree distribution.

Thus, for any d and time t, we find the node $i(d)$ such that $d_{i(d)}(t) = d$. The nodes that have degree of less than d are then those born after $i(d)$. The resulting cumulative distribution function is then $F_t(d) = 1 - \frac{i(d)}{t}$.

[4] For one such analysis, see Benaïm and Weibull [50].

Applying this technique to this random network process, we solve for $i(d)$ such that

$$d = m + m \log \left(\frac{t}{i(d)} \right).$$

This implies that

$$\frac{i(d)}{t} = e^{-\frac{d-m}{m}}$$

Thus, such a network would have distribution function described by

$$F_t(d) = 1 - e^{-\frac{d-m}{m}}$$

This is a negative exponential distribution with support from m to infinity and a mean degree of $2m$ (as it intuitively should be, as each link involves two nodes and each new node brings m links with it).

Note that this matches what we found in Section **??**, where we worked with the discrete time system, rather than the continuous time approximation. Once we are working with a mean-field approximation, the continuous time approximation is relatively minor, smoothing things out and allowing us to use differential equations, which can substantially simplify calculations. The main approximation to worry about is working with expected rather than realized values.

5.2 Preferential Attachment

With some understanding of techniques for tackling growing processes, let us enrich the process a bit.

As discussed in Section **??**, degree distributions of a number of observed networks exhibit "fat tails." Price [521], [522] pioneered the study of power distributions in networks. He adapted ideas of Simon [561] to scale-free degree distributions in a setting of growing networks. His empirical focus was on networks of citations among scientific papers. His idea was that an article would gain citations over time in a manner proportional to the number of citations the paper already had. This goes along with a story where researchers randomly find some article (e.g., via searching for key words on the internet) and then search for additional papers by tracing through the references of the first article. The more citations an article has, the larger the likelihood that it will be found and cited again. So, ignoring other issues guiding citation decisions,[5] the

[5] We will come back to discuss this in Section **??**.

probability that an article gets cited is proportional to the number of citations it already has. In the recent literature, such a link formation process was named "preferential attachment" by Barabasi and Albert [40].[6]

It is important to note that fat-tailed distributions have been found in a wide variety of applications, and so the basic ideas behind generating such distributions has a long history. Some of the first work on this was by Pareto [501], for whom the canonical power distribution is named. In the 1890's, Pareto was looking at wealth distributions across a society, and noticed that the distribution had scale-free features, where there were many more individuals who had large or small amounts of wealth than would appear in a normal or other purely random distribution. Such features were also observed in the frequency of word usage by Zipf [639] and city sizes (also known as "Zipf's law").[] Explanations for why systems should exhibit such a distribution were first put forth by Yule in 1925 [636] and Simon in 1955 [561]. Most processes that generate scale-free distributions are essentially (and sometimes unknowingly) variations on the ideas first formalized by Simon, with roots tracing back to Yule. The two basic ingredients that lead to scale-free distributions are (i) that the system grows over time so that new objects continue to enter (e.g., nodes in network applications, people in wealth applications, and cities in city-size applications) and (ii) that existing objects grow at rates that are proportional to their size. This second feature has the effect that the rich get richer (faster than the poor), and is essential to obtaining such a distribution. This proportional growth feature is also central to a close cousin of scale-free distributions: lognormal distributions. It is the specifics of the growth of the system that results in the scale-free features rather than the lognormal nature of the distribution.

Let us explore a basic preferential-attachment model in more detail. Nodes are born over time and indexed by their date of birth $i \in \{0, 1, 2..., t, ...\}$. Just as in the model above, upon birth each new node forms m links with pre-existing nodes. The difference is that way in which a new node selects which existing nodes it links to.

[6] Barabasi and Albert [40] developed a similar model to that of Price [522] except it is undirected, while Price's was directed. There are a number of studies generating power or scale-free degree distributions based on variations of preferential attachment. These include, for example, Kumar et al [**?**], whose copying method is akin to preferential attachment, as well as Dorogovtsev and Mendes [191], Levene et al [**?**], and Cooper and Frieze [163]. See Newman [480] and Mitzenmacher [446] for more discussion of such processes and their development. There are also other models generating scale-free distributions by appealing directly to the fitness of nodes and proposing that links depend on fitness and that fitness has a power distribution (e.g., see Caldarelli et al [111]).

Instead of selecting m of the nodes uniformly at random, it attaches to nodes with probabilities proportional to their degrees. For example, if one existing node has three times as many links as some other existing node, then it is three times as likely to get a given link from the newborn node. Thus, the probability that an existing node i gets a new link from the newborn node at time t is m times i's degree relative to the overall degree of all existing nodes at time t, or

$$m\frac{d_i(t)}{\sum_{j=1}^{t} d_j(t)}.$$

As there are tm total links in the system at time t, it follows that $\sum_{j=1}^{t} d_j(t) = 2tm$. Therefore, the probability that node i gets a new link in period t is

$$\frac{d_i(t)}{2t}.$$

Again, there are some details to worry about in starting such a process. So, start with a pre-existing group of m nodes each connected to each other. Now we have well-defined stochastic process. Following our discussion above, we examine the "mean-field" approximation. The mean-field, continuous-time approximation of this process is described by

$$\frac{dd_i(t)}{dt} = \frac{d_i(t)}{2t},$$

with initial condition $d_i(i) = m$. This has a solution of

$$d_i(t) = m\left(\frac{t}{i}\right)^{1/2}. \tag{5.5}$$

Thus, nodes are born over time and then grow. Just as before, the degrees of nodes can be ordered by their ages, with oldest nodes being the largest. To find the fraction of nodes with degrees that exceed some given level d at some time t, we just need to identify which node is at exactly level d at time t, and then we know that all nodes born before then are the nodes that are larger. Let $i_t(d)$ be the node which has degree d at time t, or such that $d_{i_t(d)}(t) = d$. From (5.5) it follows that

$$\frac{i_t(d)}{t} = \left(\frac{m}{d}\right)^2.$$

The fraction of nodes that have degree smaller than d at time t is the proportion born after node $i_t(d) = t\left(\frac{m}{d}\right)^2$. At time t, this is a fraction of $\frac{m^2}{d^2}$. Thus, the distribution function is

$$F_t(d) = 1 - m^2 d^{-2}.$$

This has a corresponding density or frequency distribution (for $d \geq m$) of

$$f_t(d) = 2m^2 d^{-3}.$$

Thus, the degree distribution (of expected degrees) is a power distribution with an exponent of -3.

This has the same time independence that we saw with the exponential distribution coming from a growing random network where new links were formed to existing nodes uniformly at random. This again follows since the relative degrees of nodes are determined by their relative birth dates.

To understand why this came out as an exponent of -3, and not some other exponent, let us examine the growth process. Recall from equation (5.5) that the degree of a node i as a function of time can be written as $d_i(t) = m\left(\frac{t}{i}\right)^{1/2}$. So, nodes are growing over time at a rate that is proportional to the square root of the time measured relative to their birthdate. It is this square root that translates into the -2 in the distribution function (and then the -3 in the density or frequency distribution). So, why is it a square root here? It was a particular aspect of process that led nodes to grow at a specific rate. To get a better feeling for this, suppose a node's degree grows at a rate of $\frac{dd_i(t)}{dt} = \frac{d_i(t)}{\gamma t}$. Then, following the same steps as we did before a node's degree would be $d_i(t) = m\left(\frac{t}{i}\right)^{1/\gamma}$ and the degree distribution would be described by $F_t(d) = 1 - m^\gamma d^{-\gamma}$, or a freqency of

$$f_t(d) = \gamma m^\gamma d^{-\gamma-1}.$$

A slower growth rate of any given node's degree over time (corresponding to a higher γ), leads to a distribution of degrees with a steeper fall-off in its frequency. That is, degrees get relatively more bunched at lower levels as we increase γ. How would one interpret γ? In the model we examined, m links were formed at each date and to nodes in proportion to their relative degrees. So, we had a sum of degrees of $2mt$ in the network at time t, and a chance $md_i/(2mt)$ of a node i getting a new link in period t. The γ was 2.

How could one justify a probability of $md_i/(\gamma mt)$ for node i getting a new link in period t? The γ here is not an easy parameter to justify altering. One possibility is as follows. Suppose that instead of a single node being born at time t, a group of new nodes comes in at time t. They form a fraction of links among themselves and the remaining fraction to existing links. So, for instance if they form a fraction α of links to existing nodes and $1 - \alpha$ amongst themselves, then the probability of node i (born before t) getting a new link in period t would be $\alpha md_i/(2mt)$. Therefore, $\gamma = 2/\alpha$.

So, if $\alpha = 1$ and new nodes form all of their links with pre-existing nodes then this is exactly the preferential attachment model we saw before with $\gamma = 2$. As α decreases, and new nodes form more of their link amongst themselves, then γ increases, which corresponds to slower growth in the degrees of pre-existing nodes, and a distribution with more concentration on relatively lower degrees.

In terms of an approximation, we have again looked at expectations and a continuous time approximation. Here there is an added complication. In the case where links were uniformly random, as in Section **??**, the realization of links was independent across time. Here, if a node happens to get more links at an early stage, that can snowball into more links at later stages. Moreover, this is a nonlinear effect. Thus, it is not only that it is difficult to see whether the approximation above is a good approximation for the degree distribution, but in fact it is not even clear that it is a good approximation of the distribution of expected degrees. The fact that a preferential attachment process does lead to the stated degree distribution has been verified by Bollabas, Riordan, Spencer, and Tusnady, [**?**], but by using an approach to keep track of the degree distribution directly (explicitly keeping track of the possible degree sequences that we could see emerging over time), rather than showing that mean-field approximations are accurate.

To get some feeling for such a network, and how it might differ from the previous random graph models, consider the following figure of a 25 node network which was generated using such a preferential-attachment process where each new node forms two links. To get this process started, Node 1 formed no links at birth, Node 2 formed only a link to 1, and the process was well defined from then on.

This network looks very different from the earlier models which had approximately the same average degree (Figures **??** and **??**). As the nodes are indexed by their birth dates, we see that the older nodes tend to have much higher degrees. For instance, node 2 has degree 11 while nodes 22 to 25 have only degree 2, or the links they form at birth. In the mean-field approximation, this is taken to an extreme where older nodes always have higher degrees. Even observed networks that display higher degrees for older nodes (e.g., citation networks, co-authorship networks, etc.) have degree distributions that are not so purely age dependent. Some older nodes might have few links, and some young ones might have more links. By adding a weighting parameter or "fitness," it is easy to extend the model so that some younger nodes can overtake older nodes because they are more attractive to link to, as outlined by Bianconi and Barabasi **??**. Just to get a feeling for that, suppose one changes the probability of a node getting a new link

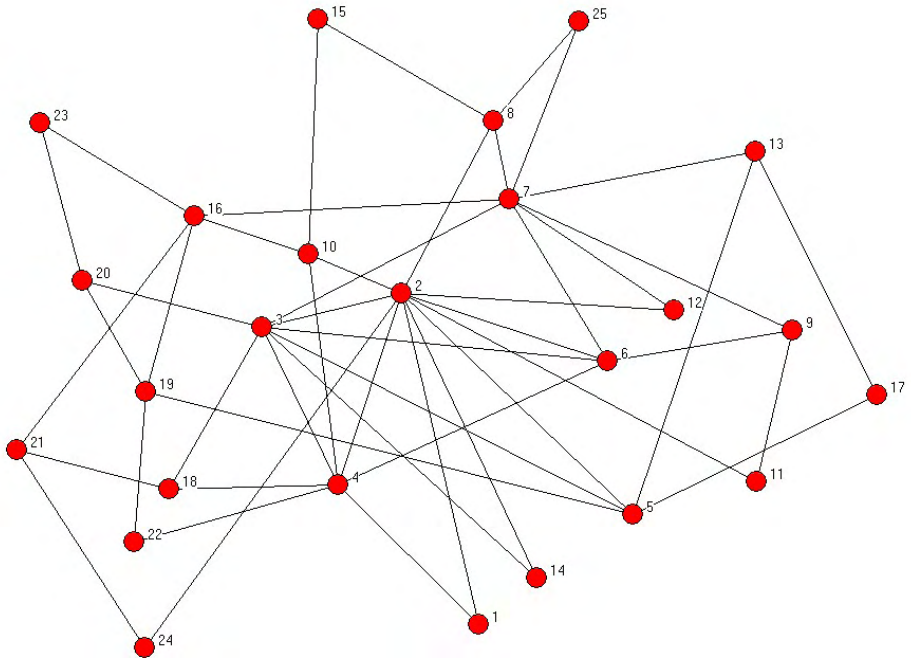

Figure 5.2. A Network of 25 nodes Formed by Preferential Attachment

via preferential attachment from $d_i/(2t)$ to $d_i v_i/(2tv)$, where v_i is a node's inherent attractiveness or fitness, and v is the average fitness in the population. Here a node can grow more quickly because of having a large d_i or a large v_i.

5.3 Hybrid Models

While network formation via preferential attachment leads to a degree distribution that is scale-free and is consistent with a power-law, many observed degree distributions match neither the exponential process we saw in Section **??** nor the preferential attachment process we saw in Section **??**. For example, consider the following degree distribution from the co-authorship network that was pictured in Figure **??**.

Here we see a degree distribution that lies somewhere between the two extremes of uniformly random link formation and preferential attachment. This suggests that a more general network formation model is needed to match observed degree distributions.

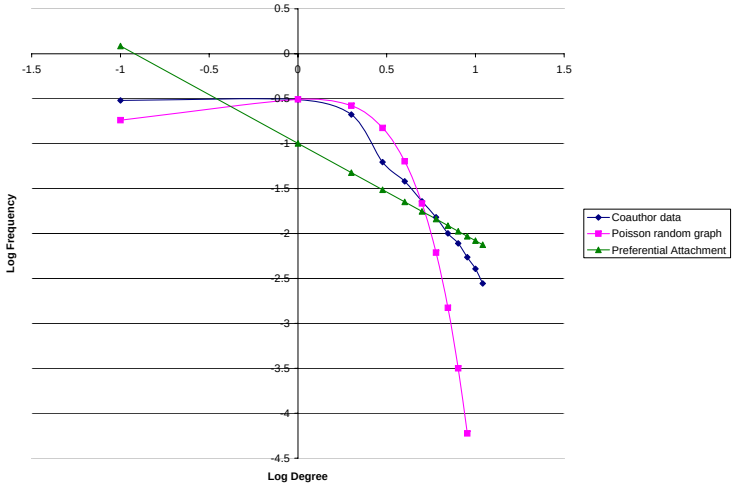

Figure 5.3. A Degree Distribution of A Co-Authorship Network that Fits Between a Uniformly Random Network and One formed via Preferential Attachment

5.3.1 Mean-Field Analyses of Growing Network Processes

There is a variety of models that are hybrids of random and preferential attachment (e.g., Kleinberg et al [**?**], Kumar et al [397] , Dorogovtsev and Mendes [191], Levene et al [**?**], and Cooper and Frieze [163], Vazquez [603], Pennock et al [513], and Jackson and Rogers [337]). Interestingly, most of these ignore the fact that the resulting degree distributions are not scale-free, but instead try to show that the distribution is at least approximately scale-free for large degrees. The explicit interest in examining the full aspects of the degree distribution and the interest in matching the features that are not scale-free first appears in Pennock et al [513], and is also a feature of Jackson and Rogers [337]. Let us examine this in more detail.

In order to work with more general processes, I start by noting the basic techniques underlying the mean-field analyses in the previous sections.

Consider any growing network such that nodes are indexed in the order of their birth and a node i's degree at time t can be represented as

$$d_i(t) = \phi_t(i),$$

where $\phi_t(i)$ is a decreasing function of i. The fact that ϕ_t is an decreasing function of i

indicates that younger nodes have lower degrees. It also means that ϕ_t is an invertible function, so that if we specify some degree d, then we can see which node has degree d at time t. The fact that degree increases with age means that the fraction of nodes with degree at least d are precisely those which are older than the node i which satisfies $\phi_i(t) = d$; that is, those nodes older than $\phi_t^{-1}(d)$. Thus, the degree distribution at time t is

$$F_t(d) = 1 - \frac{\phi_t^{-1}(d)}{t}. \tag{5.6}$$

So, whenever we can derive an expression for $d_i(t)$ that is decreasing in i, so that older nodes have more links, then we can easily derive the associated degree distribution.

5.3.2 Mixing Random and Preferential Attachment

So, now let us examine a hybrid model of link formation. Suppose for instance, that a newborn node meets existing nodes via two different processes, where we combine the formation of links uniformly at random with preferential attachment. Each newborn node forms m links, with a fraction of $\alpha < 1$ of them formed to existing nodes selected uniformly at random, and a fraction $1 - \alpha$ of them formed to existing nodes via preferential attachment.

In that case, the mean-field expression for the change in the degree of a node over time can be written as

$$\frac{dd_i(t)}{dt} = \frac{\alpha m}{t} + \frac{(1-\alpha)md_i(t)}{2mt} = \frac{\alpha m}{t} + \frac{(1-\alpha)d_i(t)}{2t}, \tag{5.7}$$

where the second expression has $(1-\alpha)m$ links being formed via preferential attachment and node i having a probability of $\frac{d_i(t)}{2mt}$ of receiving any one of them, and the first expression representing the chance of receiving one of the αm links being formed by picking uniformly at random from the t existing nodes.

(5.7) is a differential equation that has as its solution

$$d_i(t) = \phi_t(i) = \left(d_0 + \frac{2\alpha m}{1-\alpha}\right)\left(\frac{t}{i}\right)^{(1-\alpha)/2} - \frac{2\alpha m}{1-\alpha}, \tag{5.8}$$

Where d_0 is the initial number of links that a node has when it is born. From (5.8) we deduce that

$$\phi_t^{-1}(d) = t\left(\frac{d_0 + \frac{2\alpha m}{1-\alpha}}{d + \frac{2\alpha m}{1-\alpha}}\right)^{\frac{2}{1-\alpha}}. \tag{5.9}$$

Thus, from (5.6) and setting $d_0 = m$ we conclude that

$$F_t(d) = 1 - \left(\frac{m + \frac{2\alpha m}{1-\alpha}}{d + \frac{2\alpha m}{1-\alpha}} \right)^{2/(1-\alpha)} . \tag{5.10}$$

When $\alpha = 0$ this is the degree distribution $1 - \left(\frac{m}{d}\right)^2$ which is the power distribution that we found in the case of pure preferential attachment. When $\alpha \to 1$, then the limit is harder to see directly, but it approaches the exponential distribution of $F(d) = 1 - e^{-\frac{d-m}{m}}$ that we found in Section **??** for the model where links were formed uniformly at random. To see this, let $x = \frac{2\alpha}{1-\alpha}$ and then note that $\left(\frac{m+xm}{d+xm}\right)^x = \left(1 + \frac{m-d}{d+xm}\right)^x$, which for large x is approximately $\left(1 + \frac{m-d}{xm}\right)^x$, which tends to $e^{\frac{m-d}{m}}$.

5.3.3 Simulations as a Check on the Degree Distribution

We are again faced with the difficulty that we have not shown that the continuous-time mean-field process where nodes grow deterministically over time will match the actual distribution of degrees in a large random network. While this is a challenging open problem, we can perform a rough check that has become a standard technique when faced with such problems. That is, we can simulate the process for some parameter choices and check that the resulting degree distribution is well-approximated by the analysis above. While this is clearly not a guarantee, it provides some reassurance that the process is not too far off for some parameter values.

Simulating such a process is actually quite easy, especially to the extent that we just want to keep track of the resulting degree distribution and not the whole network. We can do this as follows. Let $D(t)$ be the sequence such that if node i has $d_i(t)$ links at time t, then the label i appears $d_i(t)$ times in the entries of the vector $D(t)$, as in the following figure:

$$D(t) = \underbrace{1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1}_{d_1(t) \text{ entries}} \quad \ldots \quad \underbrace{i, i, i, i}_{d_i(t) \text{ entries}} \quad \ldots$$

If each new node is forming m links then this is of length $2mt + m^2$, since each link counts for two different nodes, plus whatever we started the system with (say m^2 additional entries).

Now, let us consider what happens at time $t + 1$. The new node forms αm links uniformly at random, and $(1-\alpha)m$ links via preferential attachment. Let us take these to be integers. The αm links can be found simply by drawing αm numbers out of 1 to

t. The $(1-\alpha)m$ links via preferential attachment can be chosen by picking an entry out of $D(t)$ with equal weight on each entry. If we happen to form more than one link to an existing node, we can do three possible things: keep duplicate links, delete a duplicate link (and not redraw), or redraw according to the appropriate part of the process until the link is not a duplicate. The third option is most in the spirit of the network formation process, at least for cases where links keep track of whether two individuals have some social or economic relationship or not.[7] To be sure that this is fully specified, let us add the uniformly random links first (without replacement, so redrawing until we have αm links to distinct nodes) and then the preferential attachment links, again redrawing if any links are duplicated.

Below are some results from a set of such simulations.

[[fill in]]

5.3.4 Fitting Hybrid Degree Distributions to Data

The degree distribution described by (5.10) is useful for fitting to data because when α varies we span between the extremes of uniformly random attachment and preferential attachment. By estimating α from fitting the hybrid model to an observed network, we can get some feeling for how links might have been formed.

Such analyses have been carried out by Pennock et al [**?**] and Jackson and Rogers [337] for different variations on such a process. Let us examine a couple of examples to see how this works.

Here is a degree distribution from a network of amateur radio operators, where a link represents the fact that two operators had a radio conversation during a one month

[7] Although we might be tempted to proclaim that the other two options should lead, asymptotically, to the same distribution, by appealing to arguments such as those in Section **??** for the configuration model, we must be careful here. That argument worked with a pre-specified degree sequence that was already on a large number of nodes. Here, one of the main reasons for running simulations is to check how the process evolves when we are careful to keep track of the full evolution of the system and not to work with an approximation or with a limiting argument starting from some late point. It is conceivable, at least with pure preferential attachment, that some node will come to grow so large, that it completely dominates the system, having almost all of the links and thus gaining more and more links. Here if we do not redraw, this puts lower degree nodes at a relative disadvantage with regards to gaining new links, compared to a situation where we redraw and are forced to form new links to different nodes in each iteration. Thus, without a careful argument that applies to the full evolution of the system and takes into account starting conditions, we cannot be sure whether these details will matter.

period. These data were collected by Killworth and Bernard [372].

[[insert histogram, move table to an appendix]]

degree	ham radio
0	3
1	11
2	5
3	2
4	1
5	0
6	2
7	2
8	3
9	3
10	2
11	2
12	1
13	1
14	0
15	0
16	1
17	0
18	0
19	0
20	1
21	1
22	1
23	0
24	0
25	1
26	0
27	1

First, we directly calculate m. Since m is the number of new links formed each period, it is half of the added degree in each period. The overall degree is $2tm$, and so m is half of the average degree. In this network the average degree is 6.95, and so m

is roughly 3.5.

Next, we need to derive the parameter α which gives us the proportion of links that are formed uniformly at random versus via preferential attachment. Recall that the continuous time mean-field approximation to the degree distribution was described in (5.10) as

$$F(d) = 1 - \left(\frac{m + \frac{2\alpha m}{1-\alpha}}{d + \frac{2\alpha m}{1-\alpha}} \right)^{2/(1-\alpha)}.$$

This is non-linear in α. While there are different approaches to estimating α in such a situation, let us take a simple iterative least squares regression approach that provides fairly accurate estimates.[8] We can rewrite things in the following form:

$$\log\left(1 - F(d)\right) = \frac{2}{1-\alpha} \log\left(m + \frac{2\alpha m}{1-\alpha} \right) - \frac{2}{1-\alpha} \log\left(d + \frac{2\alpha m}{1-\alpha} \right). \tag{5.11}$$

If we now start with an initial guess of α, say α_0, we can regress $\log(1 - F(d))$ on $\log\left(d + \frac{2\alpha_0 m}{1-\alpha_0}\right)$ to estimate $\frac{2}{1-\alpha}$, and hence get an estimate α_1. We can either iterate this process until our estimate converges to some α^*, or simply examine a grid of values for α_0. Let us see what estimates that come out for a grid of values of α_0.

Table for ham radio operator fits:

α_0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.99	.999	.9999
α_1	-0.63	-0.40	-0.21	-0.04	0.12	0.28	0.44	0.61	0.79	0.89	0.98	0.9998

Here our estimate is that α is nearly 1, and so it appears that the ham radio operator network's degree distribution is best fit in the hybrid model by the extreme where links are formed uniformly at random.

To compare this to another application, we can also fit such a hybrid distribution to a data set of co-authorship relationships among economists from Goyal, van der Liej, and Moraga-Gonzalez [286]. This data set consists of researchers who published an article in a journal listed in EconLit during the 1990's. A link indicates that two researchers were co-authors on at least one article during this period.

[[insert histogram, move table to an appendix]]

[8] As discussed by Jackson and Rogers [337], one could hypothetically do a maximum likelihood estimation here. However, that requires deriving the probability of observing any given degree distribution as a function of the parameters of the model, which appears to be difficult analytically.

degree	co-author
0	24578
1	25078
2	17139
3	5069
4	3089
5	1853
6	1232
7	815
8	631
9	443
10	329
11	226
12	156
13	134
14	87
15	73
16	60
17	51
18	35
19	19
20	27
21	16
22	9
23	11
24	15
25	7
26	7
27	6
28	7
29	1
30	2
31	3
32	1
33	2
34	2
35	0
36	0
37	0
38	0

Table for co-author fits:

α_0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.56
α_1	0.43	0.45	0.48	0.51	0.54	0.58	0.62	0.69	0.79	0.56

So here, we estimate α to be about .56. This corresponds to a ratio of links formed uniformly at random to links formed via preferential attachment of about 1.27.[9]

5.4 Small Worlds, Clustering, and Assortativity

Let us now examine how growing random networks relate to other observed aspects of social networks.

5.4.1 Diameter

The diameter of a growing random network can be quite different than that of say a Poisson random network, to the extent that the growing network has very large degree nodes emerge which can serve as hubs and decrease distances. Generally, diameters and average path lengths are very difficult to calculate in neworks beyond Poisson random networks and variations on the configuration model, where the independence of the randomness provides somewhat of a toehold.

A result has been worked out by Bollobás and Riordan [81] for the special case of a preferential attachment network formation process where each node forms a single link (see also Reed [529]). They show that the network consists of a single component with diameter proportional to $\log(n)$ almost surely, while if more than one link is formed by each new node then the diameter is proportional to $\frac{\log(t)}{\log\log(t)}$. The proof of the following theorem is quite long and omitted.

PROPOSITION **5.4.1** *[Bollobás and Riordan [81]] In a preferential attachment model where each newborn node forms $m \geq 2$ links, as n grows the resulting network will consist of a single component with diameter proportional to $\frac{\log(n)}{\log\log(n)}$, almost surely.*

Thus, the diameter of a pure preferential attachment process is lower than that of a Poisson random network (which is proportional to $\log(n)$ when average degree is held

[9] This differs significantly from the Jackson and Rogers [337] estimates reported in Table **??** as they worked with a directed process and this is an undirected one. In their setting, the preferential attachment aspect is dependent on in-degree and they estimate in-degree from total degree to be 1/2 of the total degree. Here, the preferential attachment is based on the total degree. Dividing the degree by two changes the size of the degrees in the upper tail so that less preferential attachment is needed (roughly by a factor of four less).

constant), and as pointed out above, the intuition for this comes from the precess of high degree nodes which serve as hubs in the network.

One might conjecture that such a result provides an upper bound on more general growing network processes, where at least two links are formed via preferential attachment. The validity of such a conjecture is not clear, and one cannot simply extend the Bollobás and Riordan [81] argument, which is particular to pure preferential attachment. The difficulty is that as we alter the process, we change the mix of the degrees at any date, which then alters the probabilities of attachment and how quickly different nodes grow.

5.4.2 Positive Assortativity and Degree Correlation

As mentioned in Section **??**, a number of social networks have positive correlation in their degree distribution. Such correlation is absent in the Poisson random networks model, and generally in the configuration model by design. It emerges quite naturally in growing random networks, and the growth of the network might help explain why we see such correlation in social networks, a point first made by Krapivsky and Redner [393].[10] The following result is a variation on a result from Jackson and Rogers [337]. It shows a very strong form of correlation in that older nodes have distribution of neighbor's degrees that first order stochastically dominates that of younger nodes. This is a stronger conclusion than positive correlation in degree, showing that the whole distribution of neighbors' degrees is shifted up for higher degree nodes, and not just the mean of the distribution.

Let $F_i^t(d)$ denote the fraction of node i's neighbors at time t who have degree d or less.

PROPOSITION **5.4.2** *[Jackson and Rogers [337]] Consider a growing hybrid random network formation process as described in Section* **??**. *Under the mean-field estimate, a node i's degree is larger than a node j's degree at time t after both are born if and only if i is older than j. In that case, the estimated distribution of i's neighbors' degrees strictly first order stochastically dominates that of j's at each time $t > j$, and in particular $F_i^t(d) < F_j^t(d)$ for all $d < d_i(t)$.*

The first part of the proposition is obvious under the mean-field estimate, as degrees

[10] For an alternative explanation, relating to relevance of nodes, see Capocci, Caldarelli and De Los Rios [126].

grow deterministically and faster for higher degrees. The second part of the condition follows from the form of the distribution functions, as shown in the proof that follows.

It is worth stressing that the extent to which networks exhibit positive degree correlation, or have degree correlated with age depends on the application, and seem more typical of social networks than other forms.[11]

Proof of Proposition 5.4.2: Under the mean field approximation, by (5.8) the degree of node i at time t is described by

$$d_i(t) = \left(d_0 + \frac{2\alpha m}{1-\alpha}\right)\left(\frac{t}{i}\right)^{(1-\alpha)/2} - \frac{2\alpha m}{1-\alpha}. \tag{5.12}$$

If $d_i(t) > d_j(t)$, then since the above is decreasing in i, it must be that $i < j < t$. For $d < d_i(t)$,

$$F_i^t(d) = 1 - \frac{d_i(t^*(d,t))}{d_i(t)},$$

where $t^*(d,t)$ is the date of birth of a node that has degree d at time t; and for $d \geq d_i(t)$

$$1 - F_i^t(d) = \frac{0}{d_i(t)}.$$

So consider $d < d_j(t)$, as the result is clear for d such that $d_j(t) \leq d < d_i(t))$. It is thus enough to show that for any $i < j < t' < t$

$$\frac{d_i(t')}{d_i(t)} > \frac{d_j(t')}{d_j(t)}. \tag{5.13}$$

Then (5.12) implies that

$$\frac{d_i(t')}{d_i(t)} = \frac{\left(d_0 + \frac{2\alpha m}{1-\alpha}\right)(t')^{(1-\alpha)/2} - \left(\frac{2\alpha m}{1-\alpha}\right) i^{(1-\alpha)/2}}{\left(d_0 + \frac{2\alpha m}{1-\alpha}\right) t^{(1-\alpha)/2} - \left(\frac{2\alpha m}{1-\alpha}\right) i^{(1-\alpha)/2}}.$$

This expression is decreasing in i, which establishes the result. ∎

As is clear from the proof, the result also holds for many other growing processes, including exponential random networks (see Exercise **??**. However, the result will not hold if the rate of growth of degree accelerates substantially beyond being proportional to degree, as then older nodes would grow so fast that (5.13) would no longer hold.

[11] See, for instance, Adamic and Huberman [3] for evidence against the correlation of age and degree on the world wide web.

5.4.3 Clustering in Growing Random Networks

As discussed in Chapter 3, a distinguishing feature of observed networks is that they exhibit high clustering, as well as low diameters and and a positive correlation between neighbors' degrees. While the above hybrid model does a good job of matching observed degree distributions, and degree assortativity, it lacks some of the other properties, as do the extremes of preferential attachment and uniform random attachment. In particular, regardless of the choices of α and m, as the network grows the average and overall clustering will both converge to 0.

To get some feeling for this, consider the exponential random network model where links are formed to existing nodes uniformly at random. The only way a cluster can form is when a newborn node links to both ends of an existing link. Consider a node born at time $t+1$, and consider two of its newly formed links. What is the chance that a link will be present? Given that it picked the two nodes uniformly at random, and there are $t(t-1)/2$ such pairs to pick from and tm existing links in the network, it is a probability of $2m/(t-1)$, which converges to 0 as t grows. When we consider the hybrid model, or pure preferential attachment, this calculation is a bit more complicated, as now the attachment probability favors high degree nodes, and as we have just seen, they are more likely to be connected to each other than to lower degree nodes. Nevertheless, the process still has a clustering going to 0, as the likelihood that any two high degree nodes are connected to each other is still vanishing, as the number of high-degree nodes is growing rapidly enough (the fat tail of the power distribution). This is difficult to show, but has been established in a mean-field analysis as we shall see below.

There are several simple processes of network formation that are hybrid processes and that will tend to have some clustering in the limit. The simplest of these is described in Vazquez [603] in the context of finding web pages to link to. A new node first randomly chooses some node. Next, with a probability q it follows a link out from that node, and with probability $1-q$ jumps to a new node chosen uniformly at random. It continues this process until it has it has formed some set number of links. First, note that if $q = 0$, then this simply reduces to the growing network formed via uniformly random attachment. In the other extreme, where $q = 1$, only the first node is identified uniformly at random, while the other nodes are found via the network structure itself. Nodes that have higher degrees are more likely to be neighbors of the first node, and thus more likely to be found as the second node. This gives a sort of preferential attachment aspect to the network formation. If node i has degree d_i, then it can be found via this network-based search process if any d_i of its neighbors

are found in the first step. Having a higher degree, holding all else constant, leads to a proportionately higher chance of being found via this search process, just as in preferential attachment. Here we now see a reason for clustering to emerge, as nodes are finding other nodes to link to precisely by following the existing link structure of the network and are thus linking to neighboring nodes on a regular basis.

To work out the details, it helps to work with a directed version of this, for reasons explained below. With this in mind, consider a variation on a network formation process proposed and analyzed by Jackson and Rogers [337], for which it is easy to estimate clustering expressions for.

5.4.4 A Meetings-Based Network Formation Model

Consider the following meetings-based network formation process. Each new node some meets $m_r > 0$ nodes uniformly at random and forms directed links to them. Then, the new node randomly chooses m_n of the out-links of the first group of nodes, and follows those links and then forms links to each of the nodes it finds via these links. We start this process with some initial network in place with enough nodes so that any group of m_r nodes has at least m_n additional neighbors (but beyond this, the precise form of the initial network does not matter for asymptotic behavior). Let $m = m_r + m_n$ be the total number of out-links formed by each newborn node and let $d_i^{in}(t)$ be a node's indegree at date t. Let m, m_r, and m_n be integers. According to this process, for a large time t, an approximation is that a node has a probability of $\frac{m_r}{t}$ of being linked to uniformly at random, and a probability of $\frac{m_n d_i^{in}(t)}{mt}$ of being found through the search along links of the first group of nodes. In this second expression, i can potentially be found if any of its d_i^{in} neighbors are found in the first step, which occurs with probability $\frac{m_r d_i^{in}(t)}{t}$, and then conditional on one of its neighbors being found it is found with probability $\frac{m_n}{m_r m}$ where m_n is the number of out-links followed out of $m_r m$ total out-links to search.[12]

Thus, the expected change in the number of links at time t for a node with indegree $d_i^{in}(t)$ is

$$\frac{dd_i^{in}(t)}{dt} = \frac{m_r}{t} + \frac{m_n d_i^{in}(t)}{mt} \tag{5.14}$$

If we compare this to the hybrid process (5.7) it is as if we are working with $\alpha = \frac{m_r}{m}$ and $1 - \alpha = \frac{m_n}{m}$, *except* that we are missing a figure of 2 in the denominator of the

[12] These calculations ignore the chance of finding more than one neighbor, or being found via multiple paths, which are second order effects and of relatively negligible probability for large t.

second expression which comes from the difference between the directed and undirected version. This is a minor change and following the same steps as we did before leads to continuous time mean-field approximation of the distribution of in-degrees described by

$$F(d^{in}) = 1 - (rm)^{1+r} \left(d^{in} + rm\right)^{-(1+r)}, \tag{5.15}$$

where $r = \frac{m_r}{m_n}$, which matches (3.2).[13]

The above described processes are a good match for network formation in cases such that the network has a natural direction to it, such as in following links among web pages or locating scientific articles and then locating articles that are cited by those articles, etc. However, if we think of more purely social processes, where an individual meets the friends of a new friend, then it would be more natural to examine an undirected process, as friendships tend to be reciprocal. If we analyze an undirected version of the above process, we run into a complication. In an undirected version, the probability that a node with degree $d_i(t)$ is linked to by a new node at time t is more complicated. There is still a probability of $\frac{m_r}{t}$ of being linked to uniformly at random. It is also true that i can potentially be found if any of its $d_i(t)$ neighbors are found in the first step, which occurs with probability $\frac{m_r d_i(t)}{t}$. But then conditional on one of its neighbors being found, the probability that i ends up with a new link is more complicated. It depends on the total degrees of its neighbors rather than just the outdegrees of the neighbors. In the above process we knew the outdegrees of any neighbor of i to be m. However, in the undirected case, the degree of a neighbor of i is actually positively correlated with $d_i(t)$ and this correlation is time-varying. Now a node with a higher degree has a higher chance of one of its neighbors being found than a node with a lower degree; but then conditional on a neighbor being found, the higher degree node actually has a lower chance of being met through the network search process as its neighbors will tend to have more neighbors than the neighbors of a lower degree node. The first effect still dominates, but the second correction can be substantial.

5.4.5 Clustering

Let us examine the clustering in this meetings-based network formation model. Given the directed nature of the process, let us begin by examinining the percentage of

[13] This sets $d_0 = 0$ so that new nodes have no in-degree.

transitive triples defined in Section **??**. Recall that this measure is

$$Cl^{TT}(g) = \frac{\sum_{i;j\neq i;k\neq j} g_{ij}g_{jk}g_{ik}}{\sum_{i;j\neq i;k\neq j} g_{ij}g_{jk}}.$$

This transitive triple measure is examining directed links ij jk, and counting the fraction of these where the link ik is present.

We immediately see why this process leads to nontrivial clustering, since if node i finds j at random, and j has a directed link to k, then there is a nontrivial chance that i will find k through the search of j's out-links and then i will form a directed link to k. Thus, it is precisely because some nodes are met through meeting the "friends of friends" that we see clustering in the process. We can easily derive a lower bound for this clustering. First, note that the denominator of the fraction of transitive triples is simply the m^2 potential triples that are generated for each distinct i (counting across its outdegree m of j's and then each j's outdegree m of k's), and so at time t the denominator is tm^2. The numerator has the cardinality t times the number of situations for each i such that i connects to both j and k and then those two are linked to each other. Each any newborn node i will have at least m_n situations where i found k by following a link from j. Thus, a lower bound on Cl^{TT} is $\frac{tm_n}{tm^2} = \frac{m_n}{m^2} = \frac{1}{(r+1)m}$. This turns out to be the correct expression for the fraction of transitive triples when $r \geq 1$, but is only lower bound otherwise, as there are also possibilities that j and k are neighbors of each other and both found via the network search process when $r < 1$ (as then $m_n > m_r$. Let us explore this in more detail.

To develop clustering estimates, consider a special case of the process such that when $r \geq 1$ then at most one link is formed in each node found uniformly at random's neighborhood, and otherwise let $\frac{m_n}{m_r}$ be a positive integer and that exactly $\frac{m_n}{m_r}$ links are formed in each node found uniformly at random's neighborhood.

PROPOSITION **5.4.3** *[Jackson and Rogers [337]] Under a mean-field approximation the fraction of transitive triples, Cl^{TT}, tends to*

$$\begin{cases} \frac{1}{(r+1)m} & \text{if } r \geq 1, \text{ and} \\ \frac{(m-1)r}{m(m-1)(1+r)r - m(1-r)} & \text{if } r < 1. \end{cases}$$

Let us examine how the fraction of transitive triples behaves as a function of r, the relative weight on uniformly random versus network-based meetings. As r grows, then the fraction of transitive triples tends to 0, just as we should expect given that the model then operates almost uniformly at random and we know that clustering in

such models goes to 0. At the other extreme, as r becomes small we have to be a bit careful. There is a lower bound on r here, since there is always at least one node found uniformly at random so that the newborn can search its neighborhoods. So, in order to run r to be low, m must be large. For instance, fixing $m_r = 1$ implies that $r = 1/m_n = 1/(m-1)$. Then Cl^{TT} simplifies to $\frac{m-1}{2m}$ which tends to $\frac{1}{2}$ as m grows (and r shrinks).[14] To see this explicitly, recall from the lower-bound discussion above, that the denominator of the fraction of transitive triples is tm^2. Then the numerator is t times the number of ij, ik pairs that a typical i has that end up with a link between j and k. In the case where $m_r = 1$, the newborn i finds one other node j at random, attaches to it, and then attaches to all but one of its neighbors. So, i has $m-1$ completed triples of the form ij and jk. Then out of the $(m-1)(m-2)/2$ pairs of j''s and k's in j's neighborhood that i has linked to, those are linked at the rate that j's neighbors are linked to each other. That happens with a rate of $\frac{Cl^{TT}m^2}{m(m-1)/2}$. [15] So we end up with

$$Cl^{TT} = \frac{(m-1) + \frac{Cl^{TT}m^2}{m(m-1)/2}(m-1)(m-2)/2}{m^2},$$

This simplifies to $Cl^{TT} = \frac{m-1}{2m}$. The proof of the overall proposition proceeds similarly.

With the estimate of transitive triples in hand, then overall clustering (ignoring the direction of links) in the meetings-based network formation model can be estimated to tend to (see Exercise 5.8):[16]

$$\begin{cases} 0 & \text{if } r \leq 1, \text{ and} \\ \frac{6(r-1)}{(1+r)[3(m-1)(r-1)+4mr]} & \text{if } r > 1. \end{cases}$$

Here the expression is 0 at or below $r = 1$ and then tends to 0 again as r gets to be very large, and so clustering is only significantly positive in an intermediate range where $r > 1$ but r is not too large. The critical aspect that requires $r > 1$ is that total clustering does not account for the direction of links in the same way that transitive triples do. When $r < 1$, then very high degree nodes start to appear, as the preferential aspect of the attachment becomes prevalent. High degree nodes have large numbers

[14] This differs from the Jackson and Rogers [337] process, where they allow a probability of linking to nodes found. In that case, one can lower the probability of linking to the uniformly randomly found node, without having to let m grow. There at the extreme of preferential attachment, the fraction of transitive triples, overall clustering, and average clustering all go to 0.

[15] $Cl^{TT}m^2$ provides the total number of pairs of linked neighbors that a typical node will have, and then $m(m-1)/2$ is the number of such pairs.

[16] Average clustering is more cumbersome and the interested reader is referred to Jackson and Rogers [337] for details.

of pairs of neighbors and a vanishing fraction of them are connected to each other, because many of them are nodes that found the high degree node via different paths. These dominate the calculation. These do not end up dominating the calculation if we account for directions of links (as in the transitive triples calculation) or average across nodes so that then the effect of low clustering that is seen among the high degree nodes is offset by the non-vanishing clustering among low degree nodes who have many fewer pairs of neighbors and were the initiator of the links to a nontrivial fraction of them. This model thus also illustrates how careful one has to be in terms of which definition of clustering one uses.

Sketch of a Proof of Proposition 5.4.3: To derive the expression for $C^{TT}(g)$, we note that the denominator is tm^2, and then consider that the numerator is then t times the expected number of situations where some i has links ij and ik and it also turns out that there is a link between j and k. So, we need to find this expectation for a given i, and then divide by m^2. The situations where there is a pair of links ij and ik for which either jk or kj is present, break into three cases relative to how node i found j and k:

1. Both j and k were found at random.
2. One of j and k (say j) was found at random and the other by a network-based meeting.
3. Both j and k were found by network-based meetings.

Under 1, the probability of j and k being connected tends to 0 as t becomes large, just as in the uniformly random case. Under 2, j and k will tend to be connected if k was found through j, but not if k was found by search of some $j' \neq j$'s neighborhood. There are a total of m_n situations where k was found via j's neighborhood. Under 3, if j and k were found by the search of different nodes' neighborhoods, then the probability that they will be linked tends to 0. It is only in the case where they were found by search of the same node's neighborhood that they will have a nonvanishing probability of being linked. Under the process described above, this can only occur when $m_n \geq m_r$; and so let us examine that case. There are $\frac{m_n}{m_r}$ links formed by a new node in the neighborhood of any one of the nodes that were found uniformly at random, and there are m_r such neighborhoods, and so there are $m_r \frac{m_n}{m_r}(\frac{m_n}{m_r} - 1)/2$ such pairs in total. As the initial node and these links are independently and uniformly chosen, these potential clusters are completed with probability $\frac{Cl^{TT}m^2}{m(m-1)/2}$, since the initial node

i' has approximately $Cl^{TT}m^2$ completed triples of $m(m-1)/2$ possible pairs of outward links. This leads to approximately

$$\frac{Cl^{TT}mm_n}{m-1}\left(\frac{1}{r}-1\right) \tag{5.16}$$

completed triples from case 3 if $m_n \geq m_r$, and 0 otherwise. Summing across the three cases we expect a given newborn node to have

$$m_n + \frac{Cl^{TT}mm_n}{m-1}\left(\frac{m_n}{m_r}-1\right)$$

clusters out of m^2 possibilities if $m_n \geq m_r$, and m_n clusters otherwise. Thus,

$$Cl^{TT} = \frac{m_n}{m^2} + Cl^{TT}\frac{m_n}{m(m-1)}\left(\frac{m_n}{m_r}-1\right), \tag{5.17}$$

if $m_n \geq m_r$, and

$$Cl^{TT} = \frac{m_n}{m^2}$$

otherwise. Solving for C^{TT} in (5.17) yields the claimed expression. ∎

While a meetings-based model offers one explanation for how clustering might emerge, there are at least two other reasons for it. One is that nodes might be connected based on some cost and benefit structure, and we then expect clustering among groups of nodes that share low connection costs due to geographical or other characteristics. This is explored in Section **??**. Another is that groups of nodes might be born in waves. For example, Klemm and Eguíluz [383] have a variation of a preferential attachment model where nodes are declared either "active" or "inactive." A new node enters as "active" and then one existing active node is randomly de-activated (with a probability inversely proportional to its degree). New nodes attach to each active node. Then with a probability μ, each of these links is rewired to a random node in the population chosen according to preferential attachment. This process thus has a fixed number of "active" nodes and each entering node ends up linked to a proportion $1-\mu$ of the active nodes when they are born. This, coupled with the fact that the list of active nodes only changes by one each period, results in significant clustering.

5.5 Exercises

EXERCISE **5.1** *Growing Objects and Degree Distributions*

Suppose we start with a population of an object of size 1. Suppose also that a new object of size 1 is born at each date, and that existing objects double in size in each period. Over time, the sequence of populations as listed by their sizes will look like (1), (1,2), (1,2,4), (1,2,4,8), (1,2,4,8,16), and so forth.

Show that fraction of objects that have size less than d at date t is $log(d)/(t-1)$ for $d \in \{1, 2, 4, \ldots, 2^{t-1}\}$.

What is different between this and the preferential attachment system described in Section **??**? (How many "links" are added each period if we interpret this as a system where the sizes are degrees?)

EXERCISE **5.2** *Three types of Link Formation*

In the appendix of Jackson and Rogers [337] the following sort of growing network formation process is considered. Newborn nodes form links to existing nodes. An existing node gets links from a newborn in three different ways:

- some links are formed with a probability relative to the size of the existing node (as in preferential attachment),
- some links are formed with a probability depending on the total time that has already evolved (as in the growing variation of the purely random network), and
- some links are formed with a constant probability.

We can think of the second and third ways of forming links as different extensions of the idea of purely random Poisson networks to a growing set of nodes. The difference between these two is only in terms of how the probability of a link scales with the size of the society. In both cases, each existing node at some time has an equal chance of getting a new link from a newborn node. The difference is in terms of what the probability of a link is. Is it that we are keeping the average degree of newborn nodes constant - which necessitates a probability of link formation that decreases with the size of the society; or is it that we are holding the probability of a link between any two nodes constant - which necessitates a growing average degree. The analysis in Section **??** worked by holding average degree constant, but the other approach is also natural in some applications.

Allowing for an arbitrary combination of all three of these different methods of forming links leads to the following expression for the change in a node i's degree over time at a time t:

$$\frac{dd_i(t)}{dt} = \frac{ad_i(t)}{t} + \frac{b}{t} + c, \tag{5.18}$$

where a b, and c are scalars.

Solve for the degree distribution under a continuous time mean-field approximation of the process under the condition that $a > 0$ and either $c = 0$ or $a \neq 1$, and with an initial condition of $d_i(i) = d_0$. Note that in those cases, the solution to (5.18) is

$$d_i(t) = \phi_t(i) = \left(d_0 + \frac{b}{a} - \frac{c}{1-a}\right)\left(\frac{t}{i}\right)^a - \frac{b}{a} + \frac{ct}{1-a}. \tag{5.19}$$

EXERCISE **5.3** *Dying Links*

Consider a growing network process such that a newborn node forms m links with a portion α (with $1 > \alpha > 0$) uniformly at random and a portion $(1-\alpha)$ via preferential attachment. Also, in any given period qm links are destroyed, where $1 - \alpha \geq q \geq 0$ and the links are selected uniformly at random out of all links that exist at the end of the period. Solve for the degree distribution under a continuous time mean-field approximation.

EXERCISE **5.4** *Degree Distributions with Groups of Self-Attaching Newborn Nodes*

Suppose that newborn nodes come in groups of n in each period. Suppose that they attach a fraction f of their links uniformly at random to other newborn nodes, and a fraction $1 - f$ to older nodes via preferential attachment. Using a continuous time mean-field approximation, develop an expression for the degree distribution.

EXERCISE **5.5** *Stochastic Dominance in Hybrid Growing Network Models**

Consider the distribution function given in (**??**) corresponding the the hybrid growing random network model, which has support for degrees of m and above. Show that for any fixed α, the distribution associated with m strictly first order stochastically dominates an alternative distribution with $m' < m$. Show that for any fixed m, the distribution associated with α strictly second order stochastically dominates an alternative distribution with $\alpha' > \alpha$.

EXERCISE **5.6** *Degree Distributions with Growth in the Numbers of Newborn Nodes over Time*

The models we worked with in this chapter generally had a single node born at each point in time. The systems are generally unchanged if we had a fixed number of nodes come in at each date. However, if the number of newborn nodes grows over time, then

that changes the degree distribution. Consider an extension of the hybrid model to a where the number of nodes entering at each date grows over time. Let the number of new nodes entering at time t be gn_t, where n_t is the number of nodes at time t and $g > 0$ is a growth rate.

Derive an estimated degree distribution under a continuous time mean-field approximation.

EXERCISE **5.7** *Positive Assortativity in Exponential Growing Random Networks*

Show that the statement of Proposition 5.4.2 also holds for exponential growing random networks of the sort described in Section 5.1.

EXERCISE **5.8** *Overall Clustering in the Meeting-Based Network Formation Model**

Building from the proof of Proposition 5.4.3, show that overall clustering in the meetings-based network formation model, treating directed links as if they were undirected, tends to

$$\begin{cases} 0 & \text{if } r \leq 1, \text{ and} \\ \frac{6(r-1)}{(1+r)[3(m-1)(r-1)+4mr]} & \text{if } r > 1. \end{cases}$$

Hint: First argue that the overall clustering at time t can be approximated by

$$\frac{3m^2 C^{TT}}{m(m-1)/2 + m^2 + \frac{1}{t}\sum_i d_i(d_i-1)/2}, \tag{5.20}$$

and then calculate $\frac{1}{t}\sum_{i=1}^{t} d_i(d_i-1)/2$ under the mean-field approximation.

EXERCISE **5.9** *Fitting a Degree Distribution from a Hybrid Model*

Consider the following degree distribution for a network of friendships among prison inmates as collected by MacRae [422].

degree	number of prisoners
0	7
1	17
2	11
3	9
4	12
5	3
6	4
7	3
8	1

Using the techniques described in Section 5.3.4, fit the degree distribution described by (5.10) to the above data.

Chapter 6

Strategic Network Formation

While the random network models discussed in Chapters **??** and **??** are useful in growing large and complex networks that exhibit certain features, they are still lacking along some important dimensions. In particular, there are many settings where not only chance but also choice plays a central role in determining relationships. Social settings by definition involve sentient actors who have discretion in which relationships they form and maintain, and generally have discretion in how much effort, time, or resources that they devote to different relationships. Examples of this include trading relationships, political alliances, employer-employee relationships, marriages, professional collaborations, citations, emails, friendships, and so forth.

There are two central aspects to modeling networks from a "strategic" point of view. The first is that we must explicitly model the costs and benefits that arise from various networks. Doing this not only enables us to model how networks form in the face of individual incentives to form or sever links, but also provides well-defined measures of overall societal welfare. Thus, we not only have predictions about which networks might form, but we also have measures of which networks are "best" from society's point of view. The second aspect of modeling strategic network formation is making a prediction of how individual incentives translate into network outcomes. In this chapter I focus on an equilibrium method, and come back to discuss a variety of possible methods as well as dynamic models in Chapter **??**.

From the outset, it is important to emphasize what is or is not embodied in a "strategic" model. Individuals need not be Machiavellian and calculate their potential benefits and costs from each potential relationship. What is critical is that they have a tendency to form relationships that are (mutually) beneficial and to drop relationships that are not. The forces behind such incentives can be quite strong and can operate

with people even being aware that they are influenced in this way. The term strategic thus carries with it connotations that are not necessary to its application.

Some of the important conclusions from the literature on strategic network formation regard the comparison between the networks that form based on individual incentives and those networks that maximize overall societal welfare. There is often some disparity and generally a tension between the individual incentives and societal welfare. This is not surprising given that there are externalities present. For example, one of my trading partners in a market might care whom I choose as my other trading partners as that could affect the prices or other terms of trade that he or she ends up with, even though he or she may have little direct influence over my choice of whom else I trade with. The interesting aspect of this tension is how extensive and resilient it is. In particular, as we shall see, even if there are transfers present so that individuals can be subsidized to maintain relationships that would be in society's interest but are not in their own interests, it can still be impossible to maintain the socially efficient networks, under some reasonable restrictions on transfers.

Another important aspect of strategic models of network formation is that they provide answers as to *why* networks take particular forms, rather than just *how* they take particular forms. For example, growing random network models allowed us to trace certain aspects of networks, such as the shape of the degree distribution and clustering, to specific types of network formation, such as the extent to which nodes are formed uniformly at random versus via preferential attachment and whether new nodes are met by navigating the network or searching at random. While such analyses relate features of the network to features of the formation process, they do not provide an understanding of why people would tend towards preferential attachment in some settings and not others. The strategic approach ties explanations back to fundamental aspects of the setting. For example, as discussed below, the explanation behind the combination of high clustering and low diameter comes out of a strategic analysis which relates high clustering to low costs of connecting to nodes that are close in social or geographical distance, and low diameter to the benefits of accessing the information held by distant nodes. This brings us to a related point. In a situation where there is diffusion of information through a network, agents payoffs will depend on the access to information that they have. As a result, this will shape their incentives regarding which relationships to form or maintain, and ultimately will affect the network structure.

6.1 Pairwise Stability

In order to model network formation in a way that accounts for individual incentives, we first need to model the net payoffs or utility that each agent receives as a function of the network.

In this setting, the nodes of the network $N = \{1, \ldots, n\}$ will often be referred to as "players."

The overall benefit net of costs that a player enjoys from a network is modeled via a *utility function* or *payoff function*. That is, the payoff to a player i is represented by a function $u_i : G(N) \to I\!R$, where $u_i(g)$ represents the net benefit that i receives if network g is in place.[1]

The utility function captures all of the benefits net of costs that a given player experiences as a function of the network in place. Depending on the setting these can include very different things, such as the value of trading opportunities if this is a trading network or the value of information that might be obtained if this is a job-contact network. The extent to which the players in a network would "know" their own or other people's utility functions is very much context-dependent. What is most critical for the approach described here is that they be aware of changes in their own utility as they add or delete links, or at least react in terms of adding relationships that increase payoffs and delete relationships that decrease payoffs.

In order to capture the fact that forming a relationship or link between two players usually involves mutual consent, while severing a relationship only involves the consent of one player, we need an equilibrium or stability concept that differs from an off-the-shelf adaptation of a non-cooperative game theoretic solution such as Nash equilibrium (see Section 9.10 for a primer on game theory). Nash equilibrium-based solution concepts fail to capture the possibility that if two players each want to engage in a relationship then we should expect them to.

To get a feeling for this issue, let us consider a basic example (see Chapter 11 for a more detailed discussion). Consider just two individuals and a choice of whether or not to form a link. A natural inclination is to model this as a game where players (simultaneously) announce whether or not they wish to be linked to each other. If they both announce that they wish to form the link, then it is formed, while if either says that they do not wish to form the link does then the link is not formed. Thus,

[1] This can be viewed as a special case of a richer object called an allocation rule, as defined in Jackson and Wolinsky [343], which is described in more detail in Chapter **??**.

it takes mutual consent to form a relationship. Suppose the link is beneficial to both players. One might try to use the concept of Nash equilibrium. A Nash equilibrium is a choice of action by each player, such that no player would benefit by changing his or her action, given the actions of the other player(s). Unfortunately, that is not a very useful concept here. There are two equilibria: one where both players say they wish to form the link and it is formed, and another where both players say they do not wish to form the link and it is not formed. The second pair of actions form a Nash equilibrium since neither player has an incentive to change his or her action, given the (correct) anticipation that the other player will say that he or she does not want to form the link. This second equilibrium does not make much sense in a social setting, where we would expect the players to talk to each other and form the link if it is in their mutual interest. However, standard game theoretic concepts do not take this into account.[2] This is an indication of the fact that some standard game theoretic equilibrium notions are not well-suited for the study of network formation, as they do not properly account for the communication and coordination that is important in the formation of social relationships in networks.

A die-hard game theorist might respond that this is simply because the game has not been properly defined. We could explicitly model all of the communication that is available between the individuals, and then the actions that they might take in response, etc. While on the face of it this might seem reasonable, it is impractical for at least two reasons. One is that modeling the possible communication is very cumbersome. A game that incorporates all of the back and forth that might go on in forming a social relationship is complex, and yet all of the added complexity only captures a very simple idea: that two individuals should be able to coordinate on forming a link when it is in their mutual interest. Moreover, even once such a game is modeled, it might have multiple equilibria and need special refinements on beliefs and other aspects of equilibrium in order to make fairly obvious predictions. An alternative to all of this modeling is to directly define an equilibrium notion on networks that incorporates mutual consent.

A very simple stability concept that captures mutual consent is pairwise stability, as defined by Jackson and Wolinsky [343],[3] which we previewed in Section **??**.

[2] There are some refinements of Nash equilibrium (such as undominated Nash or trembling hand perfect equilibrium) that select the "natural" equilibrium of forming the link in this particular example, but fail to handle other examples. This is discussed in more detail in Chapter 11.

[3] This should not be confused with a similarly named concept that has been used in the "marriage market" literature following Gale and Shapley [**?**]. Although related, there are distinctions as the Gale

A network g is *pairwise stable* if

(i) for all $ij \in g$, $u_i(g) \geq u_i(g - ij)$ and $u_j(g) \geq u_j(g - ij)$, and

(ii) for all $ij \notin g$, if $u_i(g + ij) > u_i(g)$ then $u_j(g + ij) < u_j(g)$.

A network is pairwise stable if no player wants to sever a link and no two players both want to add a link. This comes in two parts in the definition. The requirement that no player wishes to delete a link that he or she is involved in implies that a player has the discretion to unilaterally terminate relationships that he or she is involved in. The second part of the definition can be stated in various ways. In order for a network to be pairwise stable, it is required that if some link is not in the network and one of the involved players would benefit from adding it, then the other player would suffer from the addition of the link. Another way to state this is that if a network g is such that the creation of some link would benefit both players involved (with at least one of them strictly benefiting), then g is not stable.

While pairwise stability is natural and easy to work with, there are limitations to the concept that deserve discussion (and are discussed at more length in Chapter 11). First, pairwise stability is a weak notion in that it only considers deviations on a single link at a time. Although this makes it easy to apply, if other sorts of deviations are viable and attractive, then pairwise stability could be too weak a concept. For instance, it could be that a player would not benefit from severing any single link but would benefit from severing several links simultaneously, and yet the network could still be pairwise stable. Second, pairwise stability considers only deviations by at most a pair of players at a time. It might be that some group of players could all be made better off by some more complicated reorganization of their links, which is not accounted for under pairwise stability. To the extent that larger groups can coordinate their actions in making changes in a network, a stronger solution concept might be needed. While this might sound artificial, such group actions can capture things like the expulsion or ostracism of an individual. In both of these regards, pairwise stability might be thought of as a necessary but not sufficient requirement for a network to be stable over time. Nevertheless, pairwise stability still turns out to be quite useful and often provides tight predictions about the set of stable networks without the need to consider richer deviations.

---

and Shapley notion allows a pair of individuals to simultaneously divorce their previous partners and marry each other. The pairwise stability notion defined on networks only considers one link at a time.

6.2 Efficient Networks

Next, let us turn our attention to the evaluation of the overall benefits that society sees from a given network. Payoffs not only provide an individual's perspective on the network, but also enable us to at least partially order networks with regards to the overall societal benefits that they generate.

6.2.1 Efficiency

Given that we have well-defined payoffs to players as a function of the network, there are two obvious and standard notions of welfare that we can apply.

One way of evaluating societal welfare is via a utilitarian principle, which is to say the "best" network is the one which maximizes the total utility of the society. This notion was referred to as "strong efficiency" by Jackson and Wolinsky [343], but I will simply refer to it as efficiency as in much of the subsequent literature.

A network g is *efficient* relative to a profile of utility functions $(u_1, \ldots, u_n)$ if $\sum_i u_i(g) \geq \sum_i u_i(g')$ for all $g' \in G(N)$.

It is clear that there will always exist at least one efficient network, given that there are only finitely many networks.

6.2.2 Pareto Efficiency

Another very standard tool used by economists for examining overall societal welfare is that of Pareto efficiency, as first defined by Pareto [501].

A network g is *Pareto efficient* relative to $(u_1, \ldots, u_n)$ if there does not exist any $g' \in G$ such that $u_i(g') \geq u_i(g)$ for all i with strict inequality for some i.

We say that one network *Pareto dominates* another if it leads to a weakly higher payoff for all individuals, and a strictly higher payoff for at least one. A network is then Pareto efficient if it is not Pareto dominated by any other network.

Pareto domination indicates unanimity in the ordering between two netwroks, and thus is a quite compelling argument in favor of the dominating network compared to the dominated network, at least from a purely welfaristic perspective. The difficulty, is of course, that such a unanimous ordering can be quite rare, and so while Pareto domination can help us rule out some networks, we are often faced with a very large set of Pareto efficient networks, and so it may not be very prescriptive or discriminating.

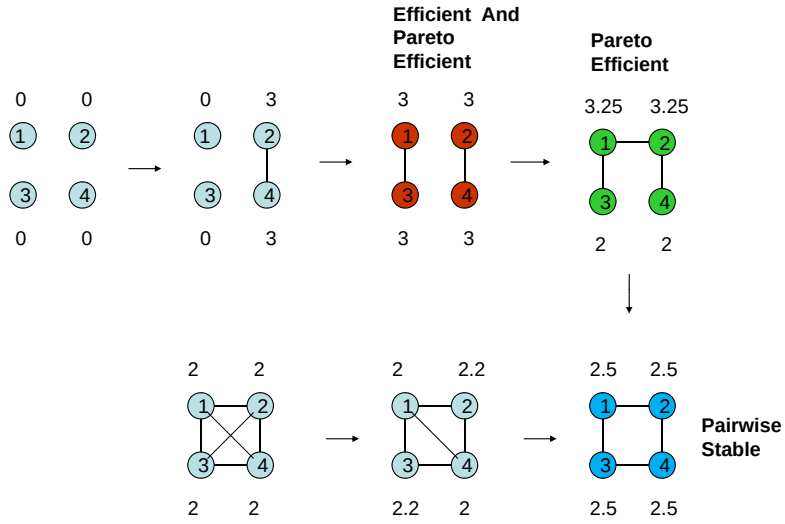

Figure 6.2.2. An Example of Efficient, Pareto Efficient, and Pairwise Stable Networks in a Four Person Society

Figure 6.2.2 illustrates these definitions in the context of a four player setting. The numbers next to the nodes are the payoffs to the respective player for each network.

There are many networks that are not pictured in Figure 6.2.2. Let any permutation of the pictured networks have correspondingly permuted payoffs to the players, and any networks that are not permutations of the pictured ones lead to payoffs of 0 for all players. The arrows in the figure indicate that the network that the arrow points away from is unstable in that some player would benefit by deleting a link, or two players would each benefit by adding a link. The network the arrow points to indicates which network would result if the player(s) who benefit from the action take the action. In this figure there is just one efficient network, marked in red, which is to have match the players into two pairs and have each player have one link. This is also a Pareto efficient network, as any other network leads to lower payoffs for some player. The efficient network here is not pairwise stable, as two disconnected players would benefit from adding a link. If such an action is taken, then the new network is the green one. The players who have formed the link have increased their payoffs from 3 to 3.25, while this has led to a lowering of the payoffs of the other two players. The green network is also Pareto efficient, as there is no other network that gives all players a weakly higher

payoff with some a strictly higher payoff. However, the green network is not pairwise stable. Here, the two players who have only one link would benefit by adding a link to each other. This then leads to the dark blue network, which is the only pairwise stable network out of those pictured. No player would gain by adding or severing a link here. We already see a conflict between stability and efficiency here, as the only pairwise stable network (or networks, if we count the permutations) is Pareto dominated by the efficient (red) network.

To better understand the relationship between efficiency and Pareto efficiency, note that if g is efficient relative to $(u_1, \ldots, u_n)$ then it must also be Pareto efficient relative to $(u_1, \ldots, u_n)$. However, the converse is not true, as we already see from Figure 6.2.2. What is true is that g is efficient relative to $(u_1, \ldots, u_n)$ if and only if is Pareto efficient relative to all payoff functions $(\widehat{u}_1, \ldots, \widehat{u}_n)$ such that $\sum_i \widehat{u}_i = \sum_i u_i$.

Thus, efficiency is a more discriminating notion and is the more natural notion in situations where there is some freedom to change the way in which utility is allocated throughout the network, for instance by reallocating value through transfers (e.g., taxes and subsidies). It can also be justified in settings where the utility functions are fixed, but where one is willing to make interpersonal comparisons of utility and take a utilitarian perspective on welfare. Pareto efficiency is a much less decisive notion, often admitting many networks, but it might be more reasonable in contexts where the payoff functions are fixed and no transfers are possible, and noting the inherent difficulties in comparing utilities across individuals.

Beyond these notions of efficiency, one may want to consider others. For instance it may be that some reallocation of value is possible, but only under the constraints that the allocations are balanced on each component. Such constraints lead to definitions of constrained efficiency, as considered in Exercises **??** and **??**.

With definitions of efficiency in hand, we can start to take a longer look at the relationship between stability and efficiency of networks.

6.3 Distance-Based Utility

I begin with a generalization from Bloch and Jackson [74] of the symmetric connections model discussed in Section **??**. The basic idea is that players get utility from their direct connections and also from their indirect connections; and the utility deteriorates with the distance between individuals. So, all else held equal, being closer to another player brings higher benefits. The rate at which is happens is captured by a benefit function.

Let $b : \{1, \ldots, n-1\} \to \mathbb{R}$ denote the net benefit that a player gets from (indirect) connections as a function of the distance between the players. The *distance-based utility model* is one where a player's utility can be written as

$$u_i(g) = \sum_{j \neq i : j \in N^{n-1}(g)} b(\ell_{ij}(g)) - d_i(g)c,$$

where $\ell_{ij}(g)$ is the shortest path length between i and j. Let $b(k) > b(k+1) > 0$ for any k and $c \geq 0$. This embodies the idea that a player sees higher benefits for having a lower distance to other players.

The benefit function is fairly general, allowing the benefits to vary with distance in a wide variety of ways. It could be that being at a distance of two links rather than one is almost as beneficial as being directly connected; or it might be that the benefits fall off dramatically. This would depend on the application and what generates the benefits. The symmetric connections model is a special case of this model, where the benefits fall off exponentially with distance, so that $b(k) = \delta^k$.

This distance-based utility model has two critical aspects: similar utility functions for different players, and benefits from indirect connections that only depend on minimum path length. While these are clearly special, it is still a setting that captures some basic aspects of the costs and benefits of many social and economic situations. It also serves as a useful benchmark, so that as we add heterogeneity or benefits that depend on other aspects of the network structure, we can understand how they change the analysis.

The following proposition shows that efficient networks in the distance-based utility model share the same features as the special case of the symmetric connections model.

PROPOSITION **6.3.1** *The unique efficient network structure in the distance-based utility model is*

(i) the complete network if $b(2) < b(1) - c$,

(ii) a star encompassing all nodes if $b(1) - b(2) < c < b(1) + \frac{(n-2)}{2} b(2)$, *and*

(iii) the empty network if $b(1) + \frac{(n-2)}{2} b(2) < c$.

So, efficient networks take simple and intuitive forms in a broad class of settings. If link costs are high relative to benefits, then it does not make sense to form any links, and so the empty network is the only efficient network (iii). If link costs are sufficiently

low ($c < b(1) - b(2)$), then it makes sense to form all links as the cost of adding a link is less then the gain from shortening a a geodesic of length at least two into a path of length one, and so the unique efficient network is the complete network (i). The more interesting case arises for intermediate costs of links relative to benefits, such that the only efficient network structure is a star (ii).

Proof of Proposition 6.3.1: To see (i), note that adding a link ij cannot decrease the utility of any $k \notin \{i, j\}$, and so if the utility to i and j increases as the result of adding a link, then total utility increases. Thus, it suffices to show that adding any link benefits the two nodes involved in the link regardless of the starting network. Note that adding a link ij cannot increase distances between them and any other nodes, and it decreases the distance between i and j. Thus, adding a link between any i and j increases each of their utilities by at least $b(1) - c - b(2)$, which is greater than 0 in case (i). Thus, adding the link increases total utility and the unique efficient network is the complete network.

Next, let us verify (ii) and (iii). To connect any k nodes involves at least $k-1$ links. A star network involves exactly $k-1$ links. A star network with $k-1$ links leads to a total utility of

$$2(k-1)(b(1)-c) + (k-1)(k-2)b(2). \tag{6.1}$$

Next, note that if a component has k nodes and $m \geq k-1$ links, then the value of the direct connections due to the links is $2m(b(1)-c)$. This leaves $\frac{k(k-1)}{2} - m$ pairs of players who are at a distance of at least 2 from each other. The value of each such indirect connection is at most $b(2)$. Therefore, the overall value of the component is at most

$$2m(b(1)-c) + (k(k-1) - 2m)\, b(2). \tag{6.2}$$

The difference between (6.1) and (6.2) is

$$2(m - (k-1))(b(2) - (b(1)-c)).$$

Since $b(2) > b(1) - c$, this is greater than 0 whenever $m > k-1$. So the value can only equal the value of the star when $m = k-1$. Any network other than a star with $k-1$ links connecting k nodes leads to a total utility that is

$$2(k-1)(b(1)-c) + X,$$

where $X < (k-1)(k-2)b(2)$, since if it is not a star and has only $k-1$ links among k nodes then some of the nodes are at a distance of more than two and at most $k-1$

pairs of nodes are directly connected. Thus, if one chooses to involve k nodes and have exactly $k-1$ links, then a star is the most efficient architecture. This implies that when $b(2) > b(1) - c$, efficient networks must involve some combinations of stars and disconnected nodes.

Next, let us show that if two stars, involving $k_1 \geq 1$ and $k_2 \geq 2$ nodes respectively, each lead to nonnegative utility, then a single star among $k_1 + k_2$ nodes leads to strictly higher total utility. This follows from (6.1), noting that the total utility from a star of $k_1 + k_2$ nodes is

$$(k_1 + k_2 - 1)[2(b(1) - c) + (k_1 + k_2 - 2)b(2)],$$

which is larger than

$$(k_1 - 1)[2(b(1) - c) + (k_1 - 2)b(2)] + (k_2 - 1)[2(b(1) - c) + (k_2 - 2)b(2)],$$

when both terms in this latter expression are nonnegative.

Thus, we can conclude that if $b(2) > b(1) - c$, then an efficient network is either a star involving all nodes or an empty network. The condition differentiating between (ii) and (iii) is exactly the calculation of whether the value of a star involving all n players is positive or negative (which is given in (6.1), setting $k = n$). ∎

We can now compare the efficient networks with those that arise if players form links in a self-interested manner. The pairwise stable networks in the distance-based utility model have similar properties to those in the symmetric connections model.

PROPOSITION **6.3.2** *In the distance-based utility model :*

(i) A pairwise stable network has at most one (non-empty) component.

(ii) For $b(2) < b(1) - c$, the unique pairwise stable network is the complete network.

(iii) For $b(1) - b(2) < c < b(1)$, a star encompassing all players is pairwise stable, but for some n and parameter values in this range is not the unique pairwise stable network.

(iv) For $b(1) < c$, in any pairwise stable network each node has either no links or else at least two links (and thus every pairwise stable network is inefficient when $b(1) < c < b(1) + \frac{(n-2)}{2} b(2)$).

The proof appears as Exercise 6.2.

As one might expect, for high and low costs to links, efficient networks coincide with pairwise stable networks. Disparities occur with intermediate link costs relative to benefits. In the range of costs and benefits such that $b(1)-b(2) < c < b(1)+\frac{(n-2)}{2}b(2)$, a star involving all players is the unique efficient network architecture, but is only sometimes pairwise stable and even then not uniquely so.

Moreover, there are situations where all pairwise stable networks are *Pareto inefficient*. To see this, consider a situation where $n = 4$ and $b(1) < c < b(1) + \frac{b(2)}{2}$, and so a star network is the unique efficient structure. Here, the only pairwise stable network is the empty network. We can argue this as follows. If a player has three links, then severing one leads to a increase in payoff of $c-b(1)$. If a player has two links, there are two possibilities: the player is in a component of all the players, or the player is in a component of just three players. In the second case, severing one of the links leads to an increase in payoff of $c-b(1)$. In the first case, the player is directly connected to two players, and then at a path length of 2 to the third player. One of the two links can be severed without increasing the path length to the third player. this leads to an increase in payoff of $c - b(1)$. Thus, it must be that each player in the network has at most one link. In that case, any player who has one link would increase his or her payoff by severing the link since it does not lead to any indirect payoffs. Although the empty network is the unique pairwise stable network, *it is not even Pareto efficient.* The empty network is Pareto dominated by a line (e.g., $g = \{12, 23, 34\}$). To see this, note that under the line, the payoff to the end players (1 and 4) is $b(1)+b(2)+b(3)-c$ which is greater than 0, and to the middle two players (2 and 3) the payoff is $2b(1)+b(2)-2c$ which is also greater than 0 since $c < b(1) + \frac{b(2)}{2}$.

Thus, there exist cost ranges for the distance-based utility model (and hence the symmetric connections model) for which all pairwise stable networks are *Pareto inefficient*, and other cost ranges where all pairwise stable networks are efficient. There are also some cost ranges where some pairwise stable networks are efficient and some other pairwise stable networks are not even Pareto efficient.

6.3.1 Externalities

The inefficiency of pairwise stable networks in the distance-based utility model stems from the externalities that are present. Externalities refer to situations where the utility or payoffs to one individual are affected by the actions of others, where those actions do not directly involve the individual in question. In the distance-based utility

model, beyond one's own links, a player in this setting can have an increase in payoffs (hence the "positive") as his or her neighbors form more links or even if indirectly connected players form more links.

Let us say that there are *nonnegative externalities* under $u = (u_1, \ldots, u_n)$ if

$$u_i(g + jk) \geq u_i(g)$$

for all $i \in N$, $g \in G(N)$ and jk such that $j \neq i \neq k$. There are *positive externalities* under $u = (u_1, \ldots, u_n)$ if there are nonnegative externalities under $u = (u_1, \ldots, u_n)$ and the inequality above is strict in some instances.

It is easy to see that the distance-based utility model is one of positive externalities, as added links can only bring players closer together.

Let us say that there are *nonpositive externalities* under $u = (u_1, \ldots, u_n)$ if

$$u_i(g + jk) \leq u_i(g)$$

for all $i \in N$, $g \in G(N)$ and jk such that $j \neq i \neq k$. There are *negative externalities* under $u = (u_1, \ldots, u_n)$ if there are nonpositive externalities under $u = (u_1, \ldots, u_n)$ and the inequality above is strict in some instances.

We shall see an example of a model with negative externalities shortly.

6.3.2 Growing Networks and Inefficiency

As there can be many pairwise stable networks, even when some are efficient we might not expect that those would be the ones to arise. How can we predict which networks are likely to emerge from a multitude of pairwise stable networks? There are a variety of approaches focusing either on refining the equilibrium concept or examining some dynamic process. To get an impression of one such dynamic, let us examine a natural and intuitive process that was introduced by Alison Watts [618], and is described as follows.

Consider a random ordering over links, where at any point in time any link is as likely as any other to be identified. If the link has not yet been added to the network, and at least one of the two players involved would benefit from adding it and the other would be at least as well off given the current network in place (and not accounting for what might happen in the future), then the link is added. If the identified link has already been added, then it is deleted if either player would (myopically) benefit from its deletion, and it is kept otherwise. If this process comes to rest, then it must be at a

pairwise stable network. It is also possible for the process to cycle. (I discuss the full range of possibilities in more detail in Chapter 11.)

Based on this process, we can deduce which pairwise stable networks will be reached in the symmetric distance-based utility model. If the empty network is the only pairwise stable network, then the process will get stuck there.[4] This happens whenever $c > b(1)$, even in cases where there are nonempty networks that are strictly preferred by all players to the empty network. In cases where $b(1) - c > b(2)$, then it is clear that all links will form and the efficient complete network will be reached. More subtle cases arise where $b(2) > b(1) - c > 0$. In this range a star is the efficient network, but players are willing to add a link to players with whom they do not have any indirect connection (or have only a sufficiently distant one). In order for a star to form, it must be that the links are identified in an order that always includes some particular player (who becomes the center) until all of the possible $n-1$ links to that player have been formed. So, for instance, if the first link that pops up is ij, then the next one (other than ij) has to be of the form ik or jk. If ik is the next one that arises, then the subsequent links that are identified to be added all have to involve i until the star forms. If any other link pops up first, the star network will not be formed (this takes some proof, which is given below). As n grows, the chance that this happens is clearly going to 0.

While Alison Watts' [618] result was stated for the symmetric connections model, it extends to the symmetric distance-based utility model.

PROPOSITION **6.3.3** *Consider the symmetric distance-based utility model in the case where $b(1) - b(2) < c < b(1)$. As the number of players grows, the probability that the above described dynamic process leads to an efficient network (star) converges to 0.*

Proof of Proposition 6.3.3: First, note that if a player forms a link at some point in the process, then that player will always be linked to at least one player from that point on. This follows from the observation that in a case where $b(1) > c$ no player would ever sever a link to a player who has no other connections (nor would that player sever his or her only link).

With this observation in hand, let us show that forming a star involves specific orders of links being identified, and that the probability of such an order being realized goes to 0. Consider a star forming with some center player, without loss of generality

[4]Here it is clear that the myopic nature of the process is critical. If players anticipate further additions to the network, they may form links that are initially costly but could later lead to net benefits. For discussion of forward looking behavior, see Chapter 11.

labeled as 1. Order the other players in terms of the last time that they end up linked to 1, and without loss of generality, label them as $2, \ldots, n$.

Note that 1 will only link to n if n is not linked to any other player when they meet as otherwise 1 is already at a distance of two links to n, and $b(2) > b(1) - c$. By the first observation above, for 1 to link to n it must be that n has not met any other players before meeting 1. Similar reasoning then implies that when 1 meets $n-1$, it must be that $n-1$ has not met any other player previously. Based on this reasoning, the only way for a star to form is for some link ij to form, then ik or jk to form, and then the center player to meet each other player before any other two players meet each other. So, suppose that ij meet first. The chance that the next two players who meet and have not met before are other than ij is $(n-2)(n-3)/[n(n-1)-2]$[5] This probability goes to 1, and so the probability that the star results goes to 0. ∎

6.3.3 The Price of Anarchy and the Price of Stability

[?] [?]

Beyond simply knowing that the network formation might lead to inefficiencies, we might also be interested in the extent to which the emergent networks are inefficient. That is, the situation is somehow worse if the stable networks are "very" inefficient compared to if they are "nearly" efficient. This issue of quantifying the social inefficiency that results from selfish individuals acting in a system is not just an issue in network settings, but is critical to a variety of settings and this has become known as the "price of anarchy."[6]

To get an idea of the price of anarchy, let us consider a special case of the distance-based utility model, where preferences are directly proportional to distance. That is, let

$$u_i(g) = \sum_{j \neq i} -\ell_{ij}(g) - d_i(g)c, \tag{6.3}$$

where ℓ_{ij} is set to ∞ if i and j are not in the same component. Such a model was considered by Fabrikant et al [216].

The way these payoffs are written they are always negative, and they can be interpreted as a sort of cost of communication.

[5] There are $(n-2)(n-3)/2$ such (unordered) pairs. The total number of possible pairs is $n(n-1)/2$, and one has already formed. So the probability is $(n-2)(n-3)/2$ divided by $[n(n-1)/2] - 1$.

[6] For example, this was studied in selfish-routing settings by Roughgarden and Tardos [542], and was named the price of anarchy by Papadimitriou [500].

Here, the *price of anarchy* is the ratio *largest* total cost (in absolute value) generated by any pairwise stable network compared to the cost of the efficient network. A ratio of 1 indicates that all pairwise stable networks are efficient, while a ratio above 1 indicates that there are higher costs (lower payoffs) associated with some pairwise stable networks than the efficient network.

We can distinguish between the best possible pairwise stable network and the worst possible pairwise stable network. This is the distinction between the "price of stability" and the "price of anarchy" (e.g., see Tardos and Wexler [585]). The *price of stability* is the ratio of the *lowest* total cost (in absolute value) generated any pairwise stable network to the cost of the efficient network. Clearly the price of anarchy will always exceed the price of stability, as one is a worst-case scenario and the other is the best-case scenario. A price of stability of 1 indicates that the efficient network will be stable, while a price of anarchy of 1 indicates that the all stable networks are efficient. A price of stability greater than 1 indicates that all stable networks are inefficient, while a price of stability of 1 and a price of anarchy greater than 1 indicates that some stable networks are efficient while others are not. These prices can differ substantially, and we can keep track of a *anarchy-stability gap*.

An easy variation on Proposition **??** shows that in this setting the unique efficient network structure is

(i) the complete network if $c < 1$, and

(ii) a star encompassing all nodes if $c > 1$.

We also see that a pairwise stable network here will have all players in one component, given that there is an infinite cost of not being connected to some other player. If $c < 1$ then it is clear that the unique pairwise (Nash) stable network is the complete network, and so there the price of anarchy and the price of stability are both 1. When $c \geq 1$, in this model a star is pairwise (Nash) stable, and so the price of stability remains 1. However, for these higher costs, there are other pairwise (Nash) stable networks and so the price of anarchy increases to be above 1. Fabrikant et al [216] provide an upper bound on this price of anarchy. The bound is fairly easy to derive by bounding the diameter of a pairwise stable network and the number of links it can contain.

PROPOSITION **6.3.4** *[Fabrikant et al [216]] The diameter of any pairwise stable network in the model described by (6.3) is at most* $2\sqrt{c}+1$ *and such a network contains at most* $n-1+\frac{2n^2}{\sqrt{c}}$ *links. Thus, the price of anarchy is no more than* $17\sqrt{c}$.

Proof of Proposition 6.3.4: First, let us bound the diameter of a pairwise stable network. Suppose that the diameter is at least $2D$ and no more than $2D+1$, where D is a positive integer. If we show that D cannot exceed $\sqrt{c}$, then it follows that the diameter cannot exceed $2\sqrt{c}+1$. Consider players i and j at a maximal distance from each other, which is at least $2D$. If they link to each other, the gain in payoff for each one will be at least

$$(2D-1)+(2D-3)+\cdots+1=D^2,$$

since they reduce the distance between them from at least $2D$ to 1 and the distance from the next closest player on the path between them from $2D-1$ to 2 (a gain of $2D-3$), and so forth. Thus, given pairwise stability, they cannot gain from adding this link and so $D^2 \leq c$ and so $D \leq \sqrt{c}$ as claimed.

Next, let us argue that the number of links in a pairwise stable network is at most $n-1+\frac{3n^2}{\sqrt{c}}$. First, there are at most $n-1$ edges which are bridges (so that the network would have more than one component if the link were removed).[7] So, we need to argue that there are at most $\frac{3n^2}{\sqrt{c}}$ links that are not bridges. Consider a link ij that is not a bridge. Let A_{ij} be the set of nodes (including j) for which the shortest path to i goes through j. If ij is deleted, the distance between i and a node in A_{ij} can become at most double the diameter of the network (denoted by d). Thus, deleting the link increases the distance costs by at most $2d|A_{ij}|$ and so this must be at least c as otherwise i should sever the link, Thus $|A_{ij}| \geq \frac{c}{2d}$. This implies that any given node i can have at most $2dn/c$ non-bridge links, so there are at most dn^2/c non-bridge links in total. Since $d \leq 3\sqrt{c}$, it follows that there are at most $3n^2/\sqrt{c}$ non-bridge links.

Finally, to derive the price of anarchy, we simply need to bound the cost of pairwise stable networks. A crude upper bound on the cost is $n(n-1)$ times the diameter plus two times c times the number of links. We also know that the cost of the efficient network when $c>1$ is that of a star, which is $2(n-1)[n-1+c]$. Therefore the price of anarchy is at most

$$\frac{(2\sqrt{c}+1)\,n(n-1)+2(n-1)c+6n^2\sqrt{c}}{2(n-1)[n-1+c]},$$

which (noting that $1<\sqrt{c}$ and $n\geq 2$) is less than $17\sqrt{c}$. ∎

This is a very loose bound, but it shows that the price of anarchy is no more than the order of the $\sqrt{c}$ in this setting, and so is the price of stability and anarchy gap.

[7] Each bridge that is removed breaks the network into one more component than we start with, and we can end up with at most n components, so there can be at most $n-1$ bridges (noting that the fact that one link is a bridge is not affected by the removal of another bridge).

In this extreme model, the price of stability is one as there is always some pairwise stable and efficient network, but there is a price of anarchy as there are some inefficient pairwise stable networks. We know more generally in the distance-based utility model that all stable networks can be inefficient.

Such price of anarchy and stability calculations are important as they provide a magnitude to the inefficiency of selfish network formation. As such calculations can be challenging outside of simple settings, the prices of stability and anarchy are still unknown for most models, especially with any heterogeneity across players.

6.4 A Co-Author Model and Negative Externalities

The analyses above for the distance-based model show us that self-centered incentives can lead to inefficient networks forming. That model has a specific form of positive externality in it: individuals benefit from indirect connections. That is, one individual can benefit because another individual has connections. The tension arising between stability and efficiency results because individuals do not account for the indirect benefits that their connections will bring to their neighbors. That is, an individual considers whether or not his or her payoff will increase when forming a link, but does not pay attention to whether the link would increase the payoffs of other players in the network.

Let us now consider another simple model of network payoffs that has a different sort of externality. Consider a situation where there are negative externalities due to links. That is, consider a situation where a given individual would rather that his or her neighbors have fewer connections rather than more. This corresponds to a situation where an individual is in competition with other indirect connections for access to the individual's neighbors. This contrasts with the connections and distance-based model where individuals draw benefits from indirect connections.[8]

This model is called the "co-author model", as introduced by Jackson and Wolinsky [343]. The story that accompanies the payoff structure is that individuals benefit from

[8] There are also models where either positive or negative externalities can result from indirect links depending on the network configuration and the players in question. This is true, for instance, if the payoff of a player is related to a centrality measure such as betweenness centrality as in Buechel and Buskens [104]. Adding a link to a network could increase some player's centrality by placing him or her between new pairs of players, even if that player is not involved in the link. So there could be positive externalities in some cases. Adding a link could also decrease some player's centrality, as it could result new paths, between other players in the network, that circumvent the given player. Thus, adding a link could exhibit negative externalities.

interacting with others, for instance, in collaborating on a research project. Beyond the benefit of having the other player put time into the project, there is also a form of synergy. The synergy is proportional to the product of times that the two researchers devote to the project. If they spend more time together, they generate more synergy. This leads to the negative externality. If an individual's collaborator increases the time spent on other projects, then the individual sees less synergy with that collaborator. Effectively, each player has a fixed amount of time to spend on projects and the time that researcher i spends on a given project is inversely related to the number of projects, $d_i(g)$, that he or she is involved in. The synergy between two researchers depends on how much time they spend together, and is captured by a term $\frac{1}{d_i(g)d_j(g)}$. Here the more projects a researcher is involved with, the lower the synergy that is obtained per project. Player i's payoff is represented by

$$u_i(g) = \sum_{j:ij\in g} \left(\frac{1}{d_i(g)} + \frac{1}{d_j(g)} + \frac{1}{d_i(g)d_j(g)} \right)$$

for $d_i(g) > 0$, and $u_i(g) = 1$ if $d_i(g) = 0$. So, the value generated by any given research project is proportional to the sum of the time that i puts into the project, the time that j puts into it, and a synergy that is dependent on an interaction between the time that the two researchers put into the project.

Note that in the co-author model there are no directly modeled costs to links. Costs are implicit in the diluted synergy as efforts are spread among more co-authors.

PROPOSITION **6.4.1** *[Jackson and Wolinsky [343]] If n is even, then the efficient network structure consists of $n/2$ separate pairs. If a network is pairwise stable and $n \geq 4$, then it is inefficient and can be partitioned into fully intra-connected components, each of which has a different number of members. Moreover, if m is the number of members of one component of a pairwise stable network and $\widehat{m}$ is the number of members of a different component that is no larger than the first, then $m > \widehat{m}^2$.*

Proof of Proposition 6.4.1:

To verify efficiency, note that

$$\sum_{i\in N} u_i(g) = \sum_{i:d_i(g)>0} \sum_{j:ij\in g} \left[\frac{1}{d_i(g)} + \frac{1}{d_j(g)} + \frac{1}{d_i(g)d_j(g)} \right],$$

so that

$$\sum_{i\in N} u_i(g) \leq 2N + \sum_{i:d_i(g)>0} \sum_{j:ij\in g} \frac{1}{d_i(g)d_j(g)},$$

and equality can only hold if $d_i(g) > 0$ for all i. Then the result follows since $\sum_{i:d_i(g)>0} \sum_{j:ij\in g} \frac{1}{d_i(g)d_j(g)} \leq n$, with equality only if $d_i(g) = 1 = d_j(g)$ for all i and j, and $3n$ is the value of $n/2$ separate pairs.

To characterize the pairwise stable networks, consider i and j who are not linked. It follows from the formula for $u_i(g)$ that i will strictly want to link to j at a given network g if and only if

$$\frac{1}{d_j(g)+1}(1+\frac{1}{d_i(g)+1}) > [\frac{1}{d_i(g)} - \frac{1}{d_i(g)+1}] \sum_{k:k\neq j, ik\in g} \frac{1}{d_k(g)},$$

(substitute 0 on the right hand side if $d_i(g) = 0$) which simplifies to

$$\frac{d_i(g)+2}{d_j(g)+1} > \frac{1}{d_i(g)} \sum_{k:k\neq j, ik\in g} \frac{1}{d_k(g)}. \tag{6.4}$$

The following facts are then true of a pairwise stable network.

1. If $d_i(g) = d_j(g)$, then $ij \in g$.

To see 1, it is enough to show that if $d_j(g) \leq d_i(g)$, then i would benefit from linking to j. Note that if $d_j(g) \leq d_i(g)$, then $\frac{d_i(g)+2}{d_j(g)+1} > 1$ while the right hand side of (6.4) is at most 1 (the average of d_i fractions). Therefore, i benefit from linking to j.

2. If $d_h(g) \leq \max\{d_k(g)|ik \in g\}$, then i benefits from a link to h.

To see 2, let j be such that $ij \in g$ and $d_j(g) = \max\{d_k(g)|ik \in g\}$. If $d_i(g) \geq d_j(g) - 1$ then $\frac{d_i(g)+2}{d_h(g)+1} \geq 1$. If $\frac{d_i(g)+2}{d_h(g)+1} > 1$ then (6.4) clearly holds for i's link to h. If $\frac{d_i(g)+2}{d_h(g)+1} = 1$, then it must be that $d_h(g) \geq 2$ and so $d_j(g) \geq 2$. This means that the right hand side of (6.4) when calculated for adding the link h will be strictly less than 1. Thus (6.4) will hold. If $d_i(g) < d_j(g) - 1$, then $\frac{d_i(g)+1}{d_j(g)} < \frac{d_i(g)+2}{d_j(g)+1} \leq \frac{d_i(g)+2}{d_h(g)+1}$. Since $ij \in g$ and g is pairwise stable, it follows from (6.4) that

$$\frac{d_i(g)+1}{d_j(g)} \geq \frac{1}{d_i(g)-1} \sum_{k:k\neq j, ik\in g} \frac{1}{d_k(g)}.$$

Also,

$$\frac{1}{d_i(g)-1} \sum_{k:k\neq j, ik\in g} \frac{1}{d_k(g)} \geq \frac{1}{d_i(g)} \sum_{k:ik\in g} \frac{1}{d_k(g)}$$

since the extra element on the right hand side is $1/d_j(g)$ which is smaller than (or equal to) all terms in the sum. Thus $\frac{d_i(g)+2}{d_h(g)+1} > \frac{1}{d_i(g)} \sum_{k:ik\in g} \frac{1}{d_k(g)}$.

Facts 1 and 2 imply that all players with the maximal number of links are connected to each other and nobody else. [By 1, they must all be connected to each other. By

2, anyone connected to a player with a maximal number of links would like to connect to all players with no more than that number of links, and hence all those with that number of links.] Similarly, all players with the next to maximal number of links are connected to each other and nobody else, and so on.

The only thing which remains to be shown is that if m is the number of members of one (fully intra-connected) component and $\widehat{m}$ is the next largest in size, then $m > \widehat{m}^2$. Notice that for i in the next largest component not to be willing to link to j in the largest component it must be that $\frac{d_i(g)+2}{d_j(g)+1} \leq \frac{1}{d_i(g)}$ (using (6.4), since all nodes to which i is connected also have $d_i(g)$ connections). Thus $d_j(g) + 1 \geq d_i(g)(d_i(g) + 2)$. It follows that $d_j(g) > d_i(g)^2$. ∎

The co-authorship model, while very different in structure from the distance-based utility model, exhibits similar features in the sense that it has a simple structure to its efficient networks, and yet the pairwise stable networks tend to be inefficient. In both models the inefficiencies are tied to externalities, but of different sorts. In the distance-based settings, when stars are efficient the center may not have an incentive to maintain links with solitary players. The externality is that when the center forms a link it benefits other players since it brings them valuable indirect connections. The failure, or tension between efficiency and stability, is due to the fact that the indirect value that the center generates is not adequately reflected in the payoffs that the center sees from a direct connection. In the co-authorship model, by forming additional connections, a player dilutes the time he or she spends with his original partners, which harms them. Here the inefficiency of stable networks stems from the fact that (up to a point) a given player sees more benefit from adding a new link than harm in terms of dilution of value from existing partnerships, while those existing partners only see harm. In both models, social and private incentives are not aligned, but for different reasons stemming from opposite sorts of externalities.

6.5 Small Worlds in an Islands-Connections Model

Before moving on to discuss the tension between stability and efficiency more generally, let us examine one more model. This is another variation on the connections model. This model shows how some of the observed features of real-world networks, such as small-world properties, can be explained from a strategic point of view. This provides a very different perspective on why we observe small-worlds than what we saw from the random networks perspective.

The reasoning behind small worlds in this model is that high clustering stems from a distance-based cost structure. Nodes that are closer (or more similar) find it cheaper to maintain links to each other and this generates high clustering. Short overall path length then comes from the fact that if there were no short enough paths between two given nodes, then even if there were a high cost to adding a link, that link would bridge distant parts of the network and bring high benefits to that pair of nodes.

This highlights an important distinction between strategic models and purely random models. The random models can identify processes which generate certain features, but do not explain why those processes might arise. In a strategic model, the explanation for a specific characteristic of a network is instead traced back to more primitive elements such as costs and benefits from social relationships. Thus, in a sense, the strategic model can be thought of as explaining "why," whereas the random-graph models can be thought of as explaining "how". This is not to say that strategic models are "better." Each modeling technique has its strengths and weaknesses, and they are quite complementary. For instance, with random graph models it is easy to produce processes which match arbitrary given degree distributions, something which is difficult (at least analytically) with a fully strategic model. Yet strategic models allow us to evaluate networks in terms of overall welfare, and trace structure back to underlying primitives.

6.5.1 The Islands-Connections Model

Consider a simple "islands" version of a truncated version of the connections model from Jackson and Rogers [335]. There are two modifications to the symmetric connections model discussed in Section **??**. First, if the minimum path length between two players is more than D links, then they do not receive any value from each other. Next, there is a "geographic" structure to costs. That is, there are K islands, each of which has J players on it. Forming a link between players i and j costs i and j each c if they are on the same island, and C otherwise, where $C > c > 0$. So, it is cheaper to link to nearby players.

This geography provides a very simple way of introducing heterogeneity among players or nodes. It is important to emphasize that the geographic structure need not be interpreted literally, but instead can also describe differences among players in terms of social or political attributes, research interests, compatibility of R&D programs, etc. In a richer model, players would be coded by whole lists of attributes and linking costs

would depend on the vectors of attributes.[9] However, this simple formulation already captures some essential aspects of social interaction and provides substantial insight into small-worlds phenomena.

The overall utility to a player i in network g is

$$u_i(g) = \sum_{j \neq i : \ell(i,j) \leq D} \delta^{\ell(i,j)} - \sum_{j:ij \in g} c_{ij},$$

where $c_{ij} = c$ if i and j are on the same island and C otherwise.[10]

For large enough D, the truncation is irrelevant. For smaller D, truncation captures the idea that benefits fall off quite dramatically beyond some threshold connection distance. So for instance, it is impossible to ask for favors from the friend of a friend of a friend of a friend. The results in this model extend to the distance-based utility model under a suitable formulation, but the truncation makes things particularly transparent.

Recall from Chapter 3 that many social networks exhibit so-called small-world characteristics embodied by a relatively low diameter and average path length, and a high clustering (compared to an independent random network). The following proposition from Jackson and Rogers [335] shows that for suitable parameter values, an islands version of the truncated connections model exhibits small-worlds characteristics. This makes clear how costs and benefits can explain small-worlds phenomena, a point first made by Carayol and Roux [127].

PROPOSITION **6.5.1** *If $c < \delta - \delta^2$ and $C < \delta + (J-1)\delta^2$, then any network that is pairwise stable or efficient is such that*

1. *the players on any given island are completely connected to each other,*
2. *the diameter and average path length are no greater than $D+1$, and*
3. *and if $\delta - \delta^3 < C$, then a lower bound on individual, average, and overall clustering is $\frac{(J-1)(J-2)}{J^2K^2}$.*[11]

[9] See Johnson and Gilles [349], as discussed in Exercise 6.13, for an alternative geographic cost structure based on distance on a line.

[10] This cost structure is the similar that of the insiders-outsiders model of Galeotti, Goyal, and Kamphorst [254], while the benefits structure is quite different: the insiders-outsiders model has almost no decay in value to distance, while this islands model matches the truncated version of the connections model (see Jackson and Wolinsky [343]). The difference in benefit structure between the islands connections model and the insider-outsider model leads to very different conclusions regarding clustering.

[11] For the bounds on clustering it is assumed that $\delta - \delta^2 \neq C$. If $\delta - \delta^2 = C$ then there is a great deal of indifference over links, and the set of pairwise stable networks explodes.

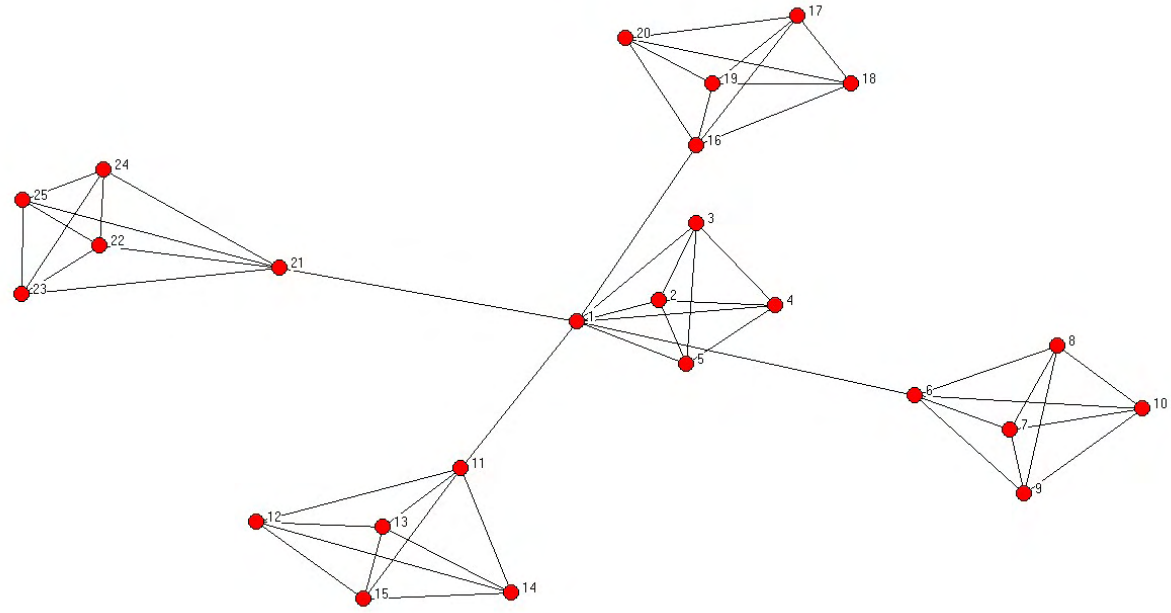

Figure 6.5.1. A Pairwise Stable "Small World" in an Islands Version of the Connections Model

The intuition behind the proposition is relatively straightforward. Low costs of connections to nearby players (those on the same island) lead to high clustering. The high value to linking to other islands (accessing many other players) leads to low average path length. The high cost to linking across islands means that there are only a few links across islands.

These properties are illustrated in Figure 6.5.1. Figure 6.5.1 is for a case where $c < .04$, $1 < C < 4.5$, $\delta = .95$. players are grouped in sets of five who are completely connected to and lie on the same island, and there are five separate islands.

This economic analysis of small worlds gives complementary insights to those of rewiring analysis of Watts and Strogatz [623] discussed in Section **??**. The random rewiring analyzed by Watts and Strogatz shows that it is possible to have both high clustering and short path length at the same time, whereas the above model gives more insight into why we should expect this to be exhibited by social networks.

Another feature distinguishing an economic modeling and a random modeling of these network characteristics concerns "shortcut" links (i.e., those which link distant parts of the network and if deleted would substantially alter the distance between the connected nodes). In a random rewiring model shortcut links would at least occasion-

ally occur in close proximity to each other. Under the strategic approach, the cost of building a second shortcut link next to an existing one would outweigh the benefit.[12]

Proof of Proposition 6.5.1: Let us first show 1. If two players on the same island are not connected in some network, then they would each gain at least $\delta - \delta^2 - c > 0$ by adding the link, and this would only help other players, and so the network cannot be pairwise stable or efficient.

Let us next show 2. Suppose that there are two players (on distinct islands), say i and j, such that $\ell(i,j) \geq D+2$. As just argued, in any pairwise stable or efficient network, j is directly connected to all members of his island and so is i. Thus i is at a distance of at least $D+1$ from each member of j's island and so enjoys no benefit from any of these players; the same is true for j from i's island. Thus, by linking to j, i would gain at least $\delta + (J-1)\delta^2 - C > 0$ (and vice versa), so this cannot be pairwise stable or efficient.

Next, let us derive a lower bound for an individual's clustering (and thus average clustering). Consider an individual with L inter-island links. All of the player's pairs of intra-island neighbors are themselves neighbors in either an efficient or pairwise stable network. Thus there are at least $(J-1)(J-2)/2$ pairs of i's neighbors that are linked out of a maximal total of $(J+L-1)(J+L-2)/2$ pairs of neighbors. This leads to a lower bound of $(J-1)(J-2)/[(J+L-1)(J+L-2)]$. Since $L \leq J(K-1)$, we have a loose lower bound of $(J-1)(J-2)/(J+J(K-1))^2$, resulting in the claimed expression.

The lower bound on the overall clustering coefficient is established as follows. For a given network, write i's clustering coefficient as a_i/b_i, where a_i is the number of links among neighbors in i's neighborhood and b_i is the number of pairs of neighbors in i's neighborhood. We have established a lower bound for a_i/b_i. Note that overall clustering is $(\sum_i a_i)/(\sum_i b_i)$ and that this is clearly greater than $\min_i(a_i/b_i)$.[13] ∎

Proposition 6.5.1 identifies small-world properties in a strong sense. The diameter is bounded above by $D+1$, and average path lengths will be smaller since each island is fully connected. Next observe that average clustering is approximated by K^{-2}. Thus, clustering can remain large when n is very large, provided that per-island population is not too small. In cases where C is large enough so that the number of inter-island

[12] I thank Yann Bramoullé for pointing this out. This does depend on the structure of the strategic model. It might be that including links which are close substitutes is valuable in changing bargaining power and payoffs. This is something discussed at more length in Section **??**.

[13] It is straightforward to check that $(a_1+a_2)/(b_1+b_2) \geq \min(a_1/b_1, a_2/b_2)$. The result then follows by induction.

links is lower (bounded by KJ), then the lower bound for clustering is even higher (on the order of $(J/(J+K))^2$); and then even for large K relative to J, the clustering is much larger than one would observe in a independent random network (which goes to 0 as the population grows, holding the probability of links constant, as discussed in Section **??**).

Proposition **??** applies to networks that are either pairwise stable or efficient, and thus it shows that there are some similarities between these sets of networks. Understanding the exact relationship between pairwise stable networks and efficient networks is complex in this model. Jackson and Rogers [335] characterize the efficient networks when the intra-island costs are low. From that they conclude that for some range of inter-island costs the pairwise stable and efficient networks coincide, whereas for other cost ranges the set of pairwise stable networks, though always exhibiting small-world features, can be quite varied. This is discussed in Exercises 6.10 and 6.11.

These results apply to cases where the intra-island cost of connections is low enough that players are completely connected within their own islands. The analysis becomes more complex when the intra-island connection cost rises, so that not all players within an island are connected. Exercise 6.12 concerns diameters in such a situation.

6.6 A General Tension Between Stability and Efficiency

In the above models, we have seen that there are settings where all pairwise stable networks are inefficient and sometimes all pairwise stable networks are even Pareto inefficient. This raises a number of interesting issues that we shall examine here, and more in Chapters **??** and **??**.

6.6.1 Transfers: Taxing and Subsidizing Links

One question is to what extent this problem can be dealt with via some transfers among the players. This might take different forms. It could be that a government or other entity intervenes to tax and subsidize different links (for example, subsidizing research and development partnerships) if it feels that there are positive externalities and individual players might form too few partnerships. It might instead be that the players themselves bargain over some payments to maintain links. For instance, the center of the star could negotiate with the other players to receive some payments

or favors for maintaining her links with the other players. In fact, intuition from the sociology literature would suggest that a player in such a central position should receive a high payoff (e.g., see Burt [105]), which could come from the implicit power that the player gets from the implicit threat of severing links, or the favors and benefits that come along with the indirect connections that the player provides. If we start to account for such reallocations, can efficiency and stability be reconciled? And, more generally, what characterizes the settings where there is a tension and when can some sort of transfers help?

Let us start with the basic question of whether or not it is possible to make some transfer payments among the players so that at least some efficient network ends up being stable. It turns out, as Jackson and Wolinsky [343] showed, that there are some very simple and natural settings where it is not possible to make transfers to align incentives and efficiency, without violating some basic principles about how transfers should or would be structured. To make this precise, we need a few definitions.

A *transfer rule* is a function $t : G \to \mathbb{R}^N$ such that $\sum_i t_i(g) = 0$ for all g.

A transfer rule can thus capture any reallocation of payoff at a given network. These payments could subsidize or tax certain links or collections of links, and could be due to intervention by some outside authority or due to bargaining by the players. What matters to players is the net payoff they receive as a function of the network. The requirement that transfers sum to 0 is usually termed a "balance" condition, and embodies the idea that the system neither depends on any outside infusion of capital in order to operate, nor does it destroy value.

In the presence of transfers, player i's net payoff becomes $u_i(g) + t_i(g)$, and this is used by the player in decisions regarding the addition or deletion of links. That is, pairwise stability is then applied where the payoffs to the players from a network g are $u_i(g) + t_i(g)$ rather than $u_i(g)$.

A first thing to note is that there is a transfer rule that aligns individual and societal incentives. That is the *egalitarian transfer rule* (denoted t^e) such that

$$u_i(g) + t_i^e(g) = \frac{\sum_j u_j(g)}{n}$$

or

$$t_i^e(g) = \frac{\sum_j u_j(g)}{n} - u_i(g).$$

This transfer rule is the one that completely equalizes all players' payoffs on any given network. Under this rule, any network that is efficient will also maximize each

individual's net payoff, as each individual equally shares in the overall societal value. While this is one way to realign individual incentives, there are reasons that such rules would not tend to arise. These are captured in the following conditions.

6.6.2 Component Balance

One condition that we would expect transfers to satisfy when they arise from a bargaining or voluntary process, and also in situations where a society worries about secession, is the following.

A transfer rule t is *component balanced* if there are no net transfers across components of the network; that is, $\sum_{i \in S} t_i(g) = 0$ for each network g and component of players $S \in \Pi(N, g)$.

Component balance requires that the value of a given component of a network is allocated to the members of that component. This is a condition that a planner or government would like to respect if they wish to avoid secession by components of the network, or if they wish to only reallocate value among the individuals who generated it.

Whether or not component balance of the transfers is a compelling condition depends on the context, and in particular on the utility functions. If the utility functions exhibit externalities across components, so that the payoffs in one component depend on how other components are organized, then it may be important to make transfers across components. This is the case, for instance, when links are cooperative ventures between firms and firms are in competition with each other. For example, if links refer to code-sharing between airlines, then a given group of airlines might care to what extent airlines whom they are not linked to are linked to each other. Applications where component balance makes more sense are those where, for instance, links represent friendships and it does not matter to a given player how players in completely separate components are organized. Component balance is also a condition that one might expect to arise naturally if the transfers are coming out of some bargaining process. For example, individuals in one component might not be willing to make transfers to another component of individuals *provided the second component's organization has no effect on the first component.*

It is important to emphasize that the result below only requires that component balance be applied in situations where there are absolutely no externalities across

components.[14] In particular, there are no externalities across components when u is component-decomposable, which is defined as follows.

A profile of utility functions u is *component-decomposable* if $u_i(g) = u_i(g|_{N_i^n(g)})$ for all i and g.[15]

6.6.3 Equal Treatment of Equals

Another basic condition regarding transfers is an equal treatment condition. The condition of equal treatment has a rich tradition in social choice (e.g., see Thomson [**?**]). It requires that two players who are completely identical according to all criteria should end up with the same transfers or allocations. It is one of the most basic fairness criteria.

In the context of social networks, the condition can be formulated as follows. Given two players i and j and a network g, let g^{ij} denote the network derived from switching the positions of i and j (and switching each of their connections).[16]

Two players i and j are *complete equals* relative to a profile of utility functions u and a network g if the following holds.

- $ik \in g$ if and only if $jk \in g$
- $u_k(\widehat{g}) = u_k(\widehat{g}^{ij})$ for all $k \notin \{i, j\}$ and for all $\widehat{g} \in G(N)$,
- $u_i(\widehat{g}) = u_j(\widehat{g}^{ij})$ and $u_j(\widehat{g}) = u_i(\widehat{g}^{ij})$ for all $\widehat{g} \in G(N)$.

Thus, two players are complete equals relative to a network and a profile of utility functions if they sit in a completely symmetric position relative to all players in the network, all other players see them as completely interchangeable in forming a network, and the two players have the same utility function as a function of the structure of the network.

A transfer rule satisfies *equal treatment of equals* relative to a profile of utility functions u if $t_i(g) = t_j(g)$ whenever i and j are complete equals relative to u and g.

---

[14] The definition here is adapted from a condition defined on allocation rules, where it can be made explicit in the actual definition that it only be applied when the allocation rule is component additive. For more on that formulation, see Chapter 12.

[15] Recall that $N_i^n(g)$ is the set of all players at a distance of no more than n from i, where n is the number of players, and so is the set of all players in i's component.

[16] Thus, $g_{kh} = g^{ij}_{kh}$ when $k \notin \{i, j\}$ and $h \notin \{i, j\}$; and $g_{jk} = g^{ij}_{ik}$ and $g_{ik} = g^{ij}_{jk}$ for all k.

Equal treatment is the weakest possible anonymity condition, stating that two players should get the same transfers when they are completely identical in terms of their position in the network and completely interchangeable in the eyes of all players, including themselves.

This condition has several justifications. From the normative side, if one is designing transfers, it captures the most basic fairness principle that one should treat identical people equally. From the positive side, in cases where one might think of the transfers arising endogenously, it captures the idea that identical people would have similar bargaining positions which would lead them to similar outcomes.

6.6.4 Incompatibility of Pairwise Stability and Efficiency

The following proposition is a variation on a result of Jackson and Wolinsky [343].[17] The proposition can be strengthened to replace efficiency with a weaker form of efficiency, as outlined in Exercise **??**.

PROPOSITION **6.6.1** *There exist component-decomposable utility functions such that every pairwise stable network relative to any component balanced transfer rule satisfying equal treatment of equals is inefficient.*

The proof is by example. It is presented for $n = 3$, but is easily adapted to any n. The utility of each player in the complete network is 4. The utility of each connected player in a linked pair is 6. The utility of players who are disconnected is 0. The efficient networks are those with two links, which have a total utility of 13, with the central player getting a utility of 4.5 and the other two players getting a utility of 4.25 each. this is pictured in Figure 6.6.4

In the absence of any transfers, the pairwise stable networks all fail to be efficient. The pairwise stable networks are only those involving a single link. In any other network some player(s) have an incentive to sever a link (every player has such an incentive in the complete network, and the center player has an incentive to do so in each of the two link networks).

Consider introducing transfers to ensure that at least one efficient network is pairwise stable. Equal treatment of equals implies that the transfers on the complete

[17] Their formulation was stated in terms of allocation rules rather than transfer rules, which is essentially equivalent (see the discussion in Chapter 12). They also required a stronger anonymity condition rather than the equal treatment condition, but their proof works with the equal treatment condition.

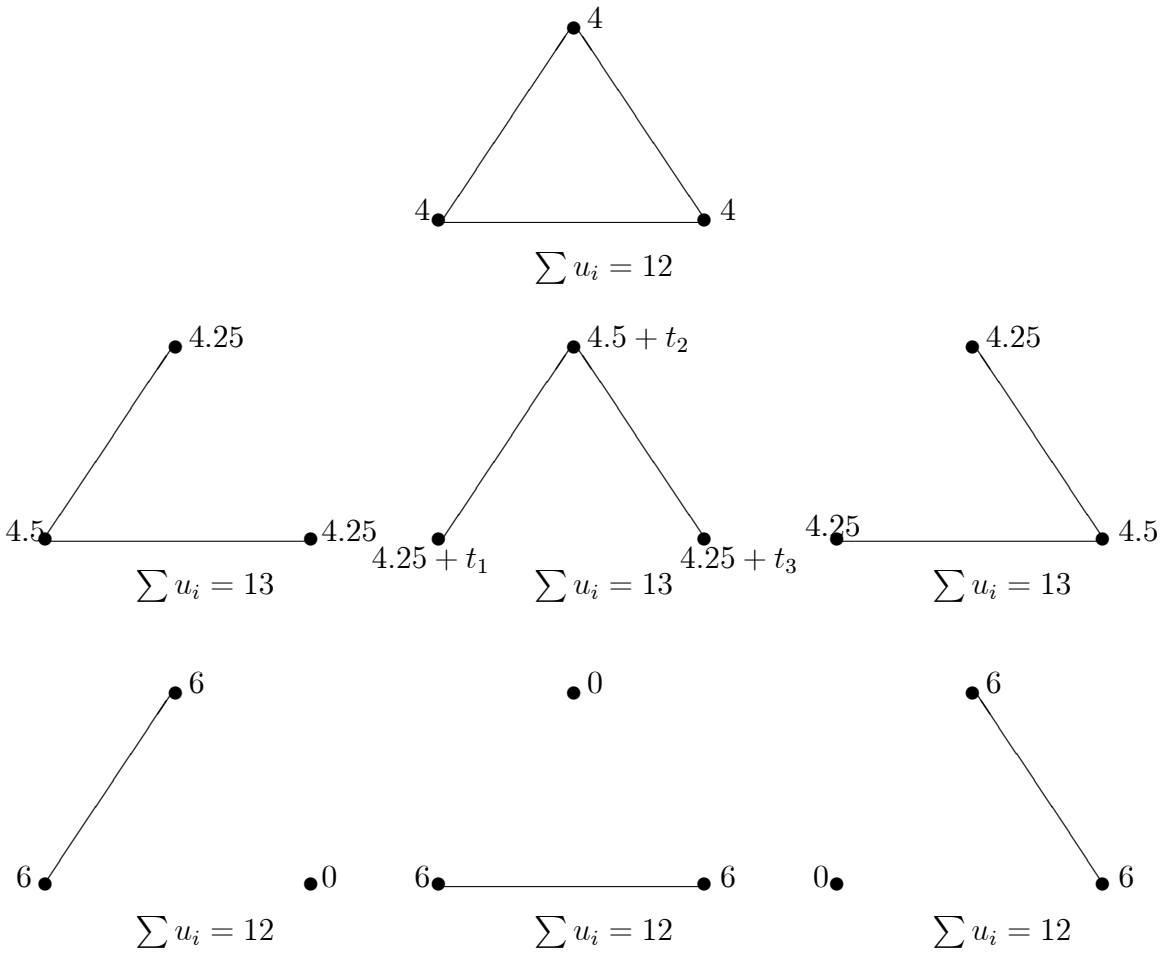

Figure 6.6.4. Payoffs such that No Pairwise Stable Network is Efficient Regardless of Transfers

network must be 0, and equal treatment plus component balance implies the same for the single link and empty networks. So, consider the middle network out of the two link networks, and let us try to introduce transfers to make it pairwise stable. Payoffs with transfers are $4.25+t_1$, $4.5+t_2$ and $4.25+t_3$, where t_1, t_2, t_3 are the transfers on this network to the three players. Given the complete symmetry between the first and third players, equal treatment implies that their transfers must be equal so that $t_1 = t_3$. For the middle two-link network to be pairwise stable, it must be that the first and third players would not gain from adding the missing link. So $t_1 = t_3 \geq -.25$. However, in order to have the network be pairwise stable we also need the second player, or center player, to be willing to keep both of the links that are in place. As that player gets a payoff of 6 if either link is deleted, it must be that $t_2 \geq 1.5$. However, now the sum of the transfers exceeds 0 and hence component balance is violated (in fact, even a basic feasibility condition is violated).

This example extends for weaker notions of efficiency and a variety of notions of stability, as for example, in Exercises 6.8 and 6.9.

While this is only an example, it has natural properties in that some intermediate sized network is efficient. It is clear that this holds for a range of examples and that this is easily modified to hold for larger societies.

Both of the conditions of component balance and equal treatment of equals are required for the result to hold. The importance of component balance is easily seen. The egalitarian transfer rule satisfied equal treatment of equals and is such that all efficient networks are pairwise stable. Individuals only have incentives to form or sever links when that increases total utility. If we drop equal treatment of equals, but keep component balance, then a careful and clever construction of transfers by Dutta and Mutuswami [200] ensures that some efficient network is strongly stable for a class of utility functions. This is stated in the following proposition, from Dutta and Mutuswami [200].

PROPOSITION **6.6.2** *[Dutta and Mutuswami [200]] If the profile of utility functions is component-decomposable and all nonempty networks generate positive total utility, then there exists a component balanced transfer rule such that some efficient network is pairwise stable. Moreover, while transfers will sometimes fail to satisfy equal treatment of equals, they can be structured so as treat completely equal players equally on at least one network that is both efficient and pairwise stable.*[18]

[18] Dutta and Mutuswami work with a variation of strong stability that is not quite a strengthening

While the details of the proof of Proposition 6.6.2 are involved, we can see some of the intuition by seeing how it would work in Example **??**. Here, we set transfers so that if player 2 is involved in a single link network, then he or she pays the other player at least 1.5. Under such transfers, the two link network with player 2 in the middle is pairwise stable, as he or she no longer gets a higher utility from severing one of the links.

Next, note that if we weaken efficiency to Pareto efficiency, Proposition 6.6.1 no longer holds. This takes a bit of proof to show generally, as one needs to work with specific transfers and then to find an algorithm to identify pairwise stable networks that are Pareto efficient. That method is described in Exercise 6.14. It is easy to see why this is true in the network in Figure **??** above, as without any transfers the single link networks are pairwise stable and Pareto efficient. Although this is of some interest, when admitting transfers Pareto efficiency is arguably not the right notion of efficiency. Instead, an efficiency notion that considers the admitted transfers is more appropriate, and then the result of Proposition 6.6.1 extends as there is an incompatibility of stability and efficiency allowing for component-balanced transfers that satisfy equal treatment of equals (as outlined in Exercise 6.8).

So, reconciling the tension between stability and efficiency requires giving something up in terms of our desired conditions of equal treatment of equals, component balance, and (constrained) efficiency; and this tension is characteristic of many network games.

There are many related questions associated with this tension that we address in the coming chapters. For instance, which settings (in terms of the structure of costs and benefits) naturally lead efficient networks to be stable? In which settings can transfers help reconcile efficiency and stability? How efficient or ineffectient are the networks that form if players bargain over payoffs during the network formation process?

6.7 Exercises

EXERCISE **6.1** *Efficiency versus Pareto Efficiency*

---

of pairwise stability, as it only considers one network to defeat another if there is a deviation by a coalition that makes all of its members strictly better off; while pairwise stability allows one of the two players adding a link to be indifferent. However, one can check that the construction of Dutta and Mutuswami extends to pairwise stability as well.

Provide an example of a society of three individuals and corresponding utility functions such that there are several Pareto efficient networks and yet only a single efficient network.

EXERCISE **6.2** *Pairwise Stability in the Distance-Based Utility Model.*

Prove Proposition 6.3.2.

EXERCISE **6.3** *Diameters in Large Pairwise Stable Networks in the Connections Model**

Consider the symmetric connections model when $c > \delta$. Show that as n grows, the diameter of any nontrivial pairwise stable network is bounded above.

Show that the same is true in the distance-based utility model for any specification of $c > b(1) > b(2) > b(3) > \cdots > b(k) > b(k+1) > \cdots$, provided that these are fixed and independent of n.

EXERCISE **6.4** *An Asymmetric Connections Model*

Consider an asymmetric version of the connections model, where players all have a common δ parameter where $0 < \delta < 1$ and the only asymmetry is that individuals might have different costs per link. In particular, suppose that each individual's cost for a link is the same for all links, so that individual i has a cost c_i for each link that i is involved with; but where it is possible that these costs differ across players.

Provide an example where the unique efficient network is not a star network, nor a complete network, nor the empty network.

Show that every efficient network has a subnetwork that is a star network (possibly only involving a subset of the players) and such that the center player in that star has a minimal cost (that is $c_i \leq c_j$ for all j).

Show that if individuals have different δ_i's, then it is possible to have an efficient network that is not empty and does not have a star network as a subnetwork.

EXERCISE **6.5** *Growing Strategic Networks*

Consider the network growth process in the distance-based utility model described in Section 6.3.2 when $n = 4$. Suppose that $b(1) - b(2) < c < b(1)$. Find the probability that an efficient network forms.

EXERCISE **6.6** *Pareto Inefficiency in the Co-Author Model.*

Show that if $n \geq 4$ is even, then any pairwise stable network in the co-author model of Section 6.4 is not only inefficient but is, in fact, Pareto dominated by some efficient network. (Hint: show that any player in any component of a pairwise stable network (including being alone) that does not involve exactly two players, receives a lower payoff than he or she would in any efficient network.)

EXERCISE **6.7** *Constrained Efficiency.*

The following notion of efficiency (from Jackson [329]) falls between efficiency and Pareto efficiency.

A network g is *constrained efficient* relative to a profile of component-based utility functions u if there does not exist any $g' \in G(N)$ and a component balanced t satisfying equal treatment of equals relative to u such that $u_i(g') + t_i(g') \geq u_i(g) + t_i(g)$ for all i with strict inequality for some i.

- Show that for any profile of component-based utility functions u, the set of efficient networks is a subset of the constrained efficient networks.

- Let $n = 5$ and consider a component based utility where all individuals are interchangeable such that the complete network generates utility of 2 per player, a component consisting of pair of individuals with one link gives each of the pair a payoff of 1, and a completely connected component among three individuals gives each player in the component a payoff of 3. All other networks generate utility of 0.

  Identify the efficient networks, and show that the completely connected network is constrained efficient but not efficient.

- Let $n = 3$. Consider a profile of utility functions u such that the complete network has a payoff of 3 for each player, any network with two links leads to a payoff of 4 to the center player and 2 to each of the other players, and all other networks lead to payoffs of 0 to all players.

  Find a Pareto efficient network relative to this u and that is not constrained efficient.

EXERCISE **6.8** *Constrained Efficiency*

Show that Proposition 6.6.1 holds when efficiency is replace by constrained efficiency.

EXERCISE **6.9** *Side Payments and the Incompatibility of Efficiency and Stability*

Consider a stronger definition of stability than pairwise stability due to Jackson and Wolinsky [343].

A network g' *defeats* a network g *allowing for side payments* if either

- $g' = g - ij$ and $u_i(g) < u_i(g')$ or $u_j(g) < u_j(g')$, or
- $g' = g + ij$ and $u_i(g') + u_j(g') > u_i(g) + u_j(g)$.

The network g is *stable relative to side payments* if it is not defeated by another network allowing for side payments.

Show that Proposition 6.6.1 holds with stability relative to side payments replacing pairwise stability, and without need for the equal treatments of equals condition.

EXERCISE **6.10** *Efficient Networks in the Islands Model*

Jackson and Rogers [335] provide the following partial characterization of efficient networks in the islands model:

Let $c < \delta - \delta^2$. In any efficient network, each island is internally completely connected and inter-island links are as follows.

1. If $C < \delta - \delta^2$ then the unique efficient network is the completely connected network.
2. If $\delta - \delta^2 < C < \delta - \delta^3$ and $K = 2$, then the efficient networks are those such that there are exactly J links between the two islands, and on at least one island each player is involved in exactly one of the J links.
3. If $\delta - \delta^3 + 2(J - L - 1)(\delta^2 - \delta^3) < C < \delta - \delta^3 + 2(J - L)(\delta^2 - \delta^3)$ and $K = 2$, then the efficient networks are those such that there are exactly $1 \leq L < J$ links between the two islands and no player is involved in more than one of these links.
4. If $\delta - \delta^3 + 2(J - 2)(\delta^2 - \delta^3) < C$ and $K = 2$, then the efficient networks have at most one link between the two islands.

Provide proof for these claims. Show also that when K is very large and $C < \delta - \delta^3$, completely connecting all players within each island, and then connecting every player on every island other than island 1 directly to the same player on island 1 can be more efficient than having intra-island links that don't all pass through the same island.

Hint for the proof of the proposition: Suppose there are just two islands and call the number of links between them L. Use the fact that

$$V(x_1, x_2, L) = 2L\delta + 2\left[Jx_1 + Jx_2 - x_1x_2\right]\left(\delta^2 - \delta^3\right) - 2L\delta^2 + 2J^2\delta^3,$$

is the utility obtained by the members of island 1 from connections to island 2 plus the reverse, where $x_i \leq L$ is the number of players on island i having links to the other island.

EXERCISE **6.11** *Inefficiency of Pairwise Stable Networks in the Islands model.*

Consider the islands model when $c < \delta - \delta^2$ and show an example with at least three islands and three players per island, such that all pairwise stable networks are nonempty and distinct from all efficient networks.

EXERCISE **6.12** *Diameter in the Islands Model.*

Consider the islands model when $\delta - \delta^2 < c < \delta$ and $C < \delta + (J-1)\delta^D$. Show that the diameter of all pairwise stable networks is no greater than $2D$.

EXERCISE **6.13** *The Spatial Connections Model*

A version of the connections model is studied by Johnson and Gilles [349] introduces geography to costs. Let there be more than three players.

The benefits take the same form as in the symmetric connection model, but there is a geography to player locations and the cost of forming a link between players i and j, c_{ij}, is related to physical distance. Let players be spaced equally on a line and i's location be at the point i, and have c_{ij} be proportional to $|i - j|$.

Provide an example where all efficient networks are nonempty, incomplete, and not stars.

Provide an example where all pairwise stable networks are ineffcient.

EXERCISE **6.14** *Pairwise Stable and Pareto Efficient Networks*

There are component balanced transfer rules, satisfying equal treatment of equals relative to utility functions that are component-based, relative to which there always exists at least one pairwise stable network that is Pareto Efficient.[19]

[19] This is a variation of a result due to Banerjee [36], based on an algorithm that is adapted to work for pairwise stability by Jackson **??**.

Given a profile of component-based utility functions u and a set of nodes $S \subset N$, let

$$g(u,S) = \text{argmax}_{g \in G(S)} \frac{\sum_{i \in S} u_i(g)}{\#N(g)}$$

denote the network with the highest per capita value to those who have links in the network out of those that can be formed by any subset of players in S. Given a component-based utility function u, find a network g^u through the following algorithm. Pick some $h_1 \in g(u,N)$ with a maximal number of links. Next, pick some $h_2 \in g(u, N \setminus N(h_1))$ with a maximal number of links. Iteratively, at stage k pick a new component $h_k \in g(u, N \setminus N(\cup_{i \leq k-1} h_i))$ with a maximal number of links. Once there are only empty networks left stop. The union of the components picked in this way defines a network g^u.

Consider component-balanced transfers such that if i and j are in the same component of g, then $u_i(g) + t_i(g) = u_j(g) + t_j(g)$.

Show that such transfers satisfy equal treatment of equals.

Show that if u is component-based, then g^u is pairwise stable and Pareto efficient (relative to the payoffs including transfers).

Show that if the algorithm above does not have the quantifier "with a maximal number of links", then the resulting g^u can fail to be pairwise stable.