communities [ncw ch. 3]
A second example is a picture of the social network of a karate club studied by Wayne Zachary and discussed in Chapter 5: a dispute between the club president and the instructor led the club to split into two clubs. Figure xsv shows the network structure with the membership in the two clubs after the division indicated by the shaded and unshaded nodes. Now a natural question is whether the structure itself contains enough information to predict the fault lines. In other words, did the split occur along a weak interface between two densely connected regions? Unlike the network in Figure xsv or in some of the earlier examples in the chapter, the two conflicting groups here are still heavily interconnected. So to identify the division in this case, we need to look for more subtle signals in the way in which edges between the groups effectively occur at lower "density" than edges within the groups. We will see that this is in fact possible, both for the definitions we consider here as well as other definitions.

A Method for Graph Partitioning

Many different approaches have been developed for the problem of graph partitioning, and for networks with clear divisions into tightly-knit regions, there is often a wide range of methods that will prove to be effective. While these methods can differ considerably in their specifics, it is useful to identify the different general styles that motivate their design.

Figure 3.13: A karate club studied by Wayne Zachary [421] — a dispute during the course of the study caused it to split into two clubs. Could the boundaries of the two clubs be predicted from the network structure?
Define the *between-ness* of a node $u$, written $\beta(u)$, as the number of shortest paths of $G$ containing $u$ - that is to say:

\[
\beta(u) = \sum_{x \in V} \varphi_x(u) \\
\varphi_x(u) = \sum_{y \in V} \left| \{ \gamma \in sp(x, y) \mid u \in \gamma \} \right|
\]
3.6. ADVANCED MATERIAL: BETWEENNESS MEASURES AND GRAPH PARTITIONING

To motivate the design of a divisive method for graph partitioning, let's think about some general principles that might lead us to remove the 7-8 edge first in Figure 3.14.

A first idea, motivated by the discussion earlier in this chapter, is that since bridges and local bridges often connect weakly interacting parts of the network, we should try removing these bridges and local bridges first. This is in fact an idea along the right lines; the problem is simply that it's not strong enough, for two reasons. First, when there are several bridges, it doesn't tell us which to remove first. As we see in Figure 3.14, where there are five bridges, certain bridges can produce more reasonable splits than others. Second, there can be graphs where no edge is even a local bridge, because every edge belongs to a triangle — and yet there is still a natural division into regions. Figure 3.15 shows a simple example, where we might want to identify nodes 1-5 and nodes 7-11 as tightly-knit regions, despite...
Figure 3.1: A network can display tightly-knit regions even when there are no bridges or local bridges along which to separate it.

However, if we think more generally about what bridges and local bridges are doing, then we can arrive at a notion that forms the central ingredient of the Girvan-Newman method. Local bridges are important because they form part of the shortest path between pairs of nodes in different parts of the network. Without a particular local bridge, paths between many pairs of nodes may have to be "reprouted" a longer way. We therefore define an abstract notion of "traffic" on the network, and look for the edges that carry the most of this traffic. Like crucial bridges and highway arteries, we might expect these edges to link different densely-connected regions, and hence be good candidates for removal in a divisive method.

We define our notion of traffic as follows. For each pair of nodes \( A \) and \( B \) in the graph that are connected by a path, we imagine having one unit of fluid "flow" along the edges from \( A \) to \( B \). If \( A \) and \( B \) belong to different connected components, then no fluid flows between them. The flow between \( A \) and \( B \) divides itself evenly along all the possible shortest paths from \( A \) to \( B \): so if there are \( k \) shortest paths from \( A \) to \( B \), then \( \frac{1}{k} \) units of flow pass along each one. We define the betweenness of an edge to be the total amount of flow it carries, counting flow between all pairs of nodes using this edge. For example, we can determine the betweenness of each edge in Figure 3.1 as follows.

- Let's first consider the 7-8 edge. For each node \( A \) in the left half of the graph and each node \( B \) in the right half of the graph, their full unit of flow passes through the 7-8 edge. On the other hand, no flow passing between pairs of nodes that both lie in the same half uses this edge. As a result, the betweenness of the 7-8 edge is \( 7 \times 7 = 9 \).
- The 5-6 edge carries the full unit of flow from each node among \( u, v, \) and \( w \) to each of the nodes in the right half of the graph, so its betweenness is \( 3 \times 3 = 9 \).

...
The Girvan-Newman Method: Successively Deleting Edges of High Betweenness. The steps of the Girvan-Newman method on the network from Figure 3.15.

Figure 3.17: The steps of the Girvan-Newman method on the network from Figure 3.15.
Write $BF(x)$ for the ranked graph induced by a breadth-first search from $x$ in $G$, where each node $y$ in $V$ is mapped to $d(x, y) \in \mathbb{N}$ its distance to $x$. Write $sp(z, y)$ for set of shortest paths from $z$ to $y$ in $G$. 
(i) Show that neighbours in $G$ are at most one rank remote in $BF(x)$.

Show that deleting all edges $(y, z)$ such that $d(x, y) = d(x, z)$ (edges of constant rank) in $BF(x)$ obtains a directed acyclic graph $BF^-(x)$.

Since, $d(x, y) + d(y, z) \leq d(x, z)$, if $d(y, z) = 1$, then $|d(x, y) - d(x, z)| \leq 1$. Directedness and acyclicity follow from the fact that all paths are now strictly rank-increasing.
Write \( ri(z, y) \) for set of paths from \( z \) to \( y \) in \( BF(x) \) with a strictly increasing rank; equivalently, the set of directed paths from \( z \) to \( y \) in \( BF^-(x) \).

Write:
- \( y^- \) for the set of immediate parents of \( y \) in \( BF^-(x) \)
- \( z^+ \) for the set of immediate successors of \( z \) in \( BF^-(x) \)

We have \( y^- = \emptyset \) iff \( y \) is the root \( x \) of \( BF^-(x) \), \( z^+ = \emptyset \) iff \( z \) is a leaf in \( BF^-(x) \).
(ii) Set $\sigma(x, y) = |ri(x, y)|$, and show that this is the number of shortest paths from $x$ to $y$.

Show that $\sigma(z, z) = 1$, and, conversely, $\sigma(z, y) > 0$, $\sigma(y, z) > 0 \Rightarrow z = y$. 

(iii) Show that $\sigma(x, y) > 0$, and $\sigma(x, z) \leq \sigma(x, y)$ if $z \in y^-$. 
(iv) Show that $\sigma$ verifies:

\[
\begin{align*}
\sigma(x, x) &= 1 \\
\sigma(x, y) &= \sum_{z \in y^-} \sigma(x, z) \quad \text{if } x \neq y
\end{align*}
\]
(vi) Define the *between-ness* of a node $u$, written $\beta(u)$, as the number of shortest paths of $G$ containing $u$ - that is to say:

$$
\beta(u) = \sum_{x \in V} \varphi_x(u)
$$

$$
\varphi_x(u) = \sum_{y \in V} \left| \{ \gamma \in sp(x, y) \mid u \in \gamma \} \right|
$$

Show:

$$
\varphi_x(u) = \sigma(x, u) + \sum_{v \in u^+} \varphi_x(v) \cdot \sigma(x, u) / \sigma(x, v)
$$
For every $x$, build $BF(x)$ and compute for each node $y$ the number $\sigma(x, y)$ using the top-down formula of subquestion (iv); then use the bottom-up formula just above, to compute $\varphi_x(u)$; to obtain $\beta(u)$, sum over $x$; the complexity is linear in the number of edges and linear in the number of nodes, so quadratic in $G$, hence cubic if one wants to pick up nodes of maximal betweenness.
is between-ness a good notion?
Do topological models provide good information about electricity infrastructure vulnerability?
Mean distance between nodes

**Connectivity loss**

\[ C = 1 - \frac{n(x,g)}{n(g)} \]

- \( n(g) \) = total number of generators
- \( n(x,g) \) = number of generators connected to \( x \)

Blackout sizes
as calculated from a model of cascading failure in a power system

\[ P_i = \sum_{j=1}^{n} \frac{(\theta_i - \theta_j)}{X_{ij}} \]
Simulated response of the IEEE 300 bus network to directed attacks

top: mean path length
middle: connectivity loss
bottom: size of blackout
random failures are averaged over 20 trials

differences between attack results and random failures

Shading ±σ for random failures
previously ...

between-ness a measure of the flow through a node (or edge)
today ...

triadic closure ...

strength of weak ties

architecture of social networks

how a network evolves over time

interpreting between-ness
3.1. TRIADIC CLOSURE

Before new edges form.

After new edges form.

Figure 3y2: If we watch a network for a longer span of time, we can see multiple edges forming — some form through triadic closure while others such as the DxG edge form even though the two endpoints have no neighbors in common. The fact that the BxC edge has the effect of "closing" the third side of this triangley If we observe snapshots of a social network at two distinct points in time, then in the later snapshot, we generally find a significant number of new edges that have formed through this triadic closure operation between two people who had a common neighbor in the earlier snapshot. Figure 3y2, for example, shows the new edges we might see from watching the network in Figure 3y1 over a longer span of time.

The Clustering Coefficient.

The basic role of triadic closure in social networks has motivated the formulation of simple social network measures to capture its prevalence. One of these is the clustering coefficient [320411]. The clustering coefficient of a node A is defined as the probability that two randomly selected friends of A are friends with each other. In other words, it is the fraction of pairs of A's friends that are connected to each other by edges. For example, the clustering coefficient of node A in Figure 3y2 is 1/6 because there is only the single CxD edge among the six pairs of friends BxC, BxD, BxE, CxD, CxE, and DxE, and it has increased to 1/2 in the second snapshot of the network in Figure 3y2. Because there are now the three edges BxC, CxD, and DxE among the same six pairs.

In general, the clustering coefficient of a node ranges from 0 when none of the node's friends are friends with each other to 1 when all of the node's friends are friends with each other, and the more strongly triadic closure is operating in the neighborhood of the node, the higher the clustering coefficient will tend to be.

Clustering Coefficient(A) = fraction of A's friends who are friends

compute cc(A) before and after
cc is genetically inherited

attach-and-introduce model

espistemo point: recreate the process behind a feature - see Miguel’s lectures; also Panconesi’s lecture on web compressibility
interpreting between-ness: where does info come from; different time scales and steady states?

philosophical stake: fine time-structure of information propagation

compute the derivative of gossip upper bounds?

distance contraction, curvature?

local bridges

no friends in common

low cc

bridges

what is propagating?

compute the derivative of gossip upper bounds?

philosophical stake: fine time-structure of information propagation
strong ties (the stronger links, corresponding to friends), and weak ties

strong triadic closedness

Figure vpy: Each edge of the social network from Figure vpw is labeled here as either a strong tie \( S \) or weak tie \( W \) to indicate the strength of the relationship. The labeling in the figure satisfies the Strong Triadic Closure Property at each node: if the node has strong ties to two neighbors, then these neighbors must have at least a weak tie between them. In other words, if we were to look at Figure vpv as it is embedded in a larger ambient social network, we would likely see a picture that looks like Figure vpwp.

Here the edge isn't the only path that connects its two endpoints; though they may not realize it, A and B are also connected by a longer path through F and G and H. This kind of structure is arguably much more common than a bridge in real social networks, and we use the following definition to capture it. We say that an edge joining two nodes A and B in a graph is a local bridge if its endpoints A and B have no friends in common—in other words, if deleting the edge would increase the distance between A and B to a value strictly more than two. We say that the span of a local bridge is the distance its endpoints would be from each other if the edge were deleted. Thus, in Figure vpwp, the A-B edge is a local bridge with span four; we can also check that no other edge in this graph is a local bridge since for every other edge in the graph, the endpoints would still be at distance two if the edge were deleted.

Notice that the definition of a local bridge already makes an implicit connection with triadic closure in that the two notions form conceptual opposites: an edge is a local bridge precisely when it does not form a side of any triangle in the graph.
3.2. THE STRENGTH OF WEAK TIES

Strong Triadic Closure says the B-C edge must exist, but the definition of a local bridge says it cannot.

Figure 1.6: If a node satisfies Strong Triadic Closure and is involved in at least two strong ties, then any local bridge it is involved in must be a weak tie. The figure illustrates the reason why: if the A-B edge is a strong tie, then there must also be an edge between B and C, meaning that the A-B edge cannot be a local bridge.

We're going to justify this claim as a mathematical statement—that is, it will follow logically from the definitions we have so far, without our having to invoke any yet-unformalized intuitions about what social networks ought to look like. In this way, it's a different kind of claim from our argument in Chapter 1 that the global friendship network likely contains a giant component. That was a thought experiment—albeit a very convincing one—requiring us to believe various empirical statements about the network of human friendships—empirical statements that could later be confirmed or refuted by collecting data on large social networks. Here, on the other hand, we've constructed a small number of specific mathematical definitions—particularly, local bridges and the Strong Triadic Closure Property—and we can now justify the claim directly from these.

The argument is actually very short, and it proceeds by contradiction. Take some network, and consider a node A that satisfies the Strong Triadic Closure Property and is involved in at least two strong ties. Now suppose A is involved in a local bridge—say, to a node B—that is a strong tie. We want to argue that this is impossible, and the crux of the argument is depicted in Figure 1.6. First, since A is involved in at least two strong ties, and the edge to B is only one of them, it must have a strong tie to some other node, which we'll call C. Now let's ask: is there an edge connecting B and C? Since the edge from A to B is a local bridge, A and B must have no friends in common, and so the B-C edge must not exist. But this contradicts Strong Triadic Closure, which says that since the A-B and local bridge on A (2 s) is weak <= STC