Chapter 8
Measuring Networks

Abstract We have adopted the view of graphs and, more generally, cell complexes as a domain upon which we may apply the tools of calculus to formulate differential equations and to analyze data. An important aspect of the discrete differential operators is that *the operators are defined by the topology of the domain itself*. Therefore, in an effort to provide a complete treatment of these differential operators, we examine in this chapter the properties of the network which may be extracted from the structure of these operators. In addition to the network properties extracted directly from the differential operators, we also review other methods for measuring the structural properties of a network. Specifically, the properties of the network that we consider are based on distances, partitioning, geometry, and topology. Our particular focus will be on the *measurement* of these properties from the graph structure. Applications will illustrate the use of these measures to predict the importance of nodes and to relate these measures to other properties of the subject being modeled by the network.

Measures of network properties have been important in several applications. Graph measures may be used to summarize node or network properties to help understand the network. These measures may also be used to make predictions about the separability of the network or the importance of individual nodes or edges. A different use of graph measures has been to predict properties of the subject which is being modeled computationally as a network. For example, a series of network measures have been used to determine whether the configuration of a network of cells inside tissue imaged in a histological section can predict cancer [101, 180]. These network measures have also been widely used in chemical graph theory to predict various structural and behavioral properties of molecules. In this chapter, we review several types of measures that describe the connectedness, topology, or geometry of a network and then give applications of these measures in social networking and chemical graph theory. Readers wishing to explore the network measurement topic further are referred to the excellent review article of Costa et al. [89].

L.J. Grady, J.R. Polimeni, *Discrete Calculus*,
DOI 10.1007/978-1-84996-290-2_8,

8.1 Measures of Graph Connectedness

In 1947 the chemist H. Wiener observed that a particular measure on the graph that represents an alkane molecule allows one to predict the boiling point of the compound. Pursuing this line of research led Wiener to show that this same distance measure could be used to calculate a series of chemical properties that had previously escaped prediction. When interest in this work was reawakened in the 1970s, it triggered a search within chemical graph theory for other distance-based "topological indices" that could be used to predict various properties of molecules (see Sect. 8.5.2 for a detailed example of this line of work). More recently, a set of similar measures has been used to describe the characteristics of complex networks. Almost all of these quantities are derived from measures of distance on a graph, which we now review.

8.1.1 Graph Distance

We first review the distance between two nodes on a weighted graph that was defined in Chap. 4. (If there are no weights specified and the graph is embedded in $\mathbb{R}^N$, then the edge weights can be set to reflect the Euclidean length of the edge.) If we consider a graph with any set of positive weights associated to the edge set, then given these edge weights we may define the distance between any pair of nodes using the distance operator $\mathcal{D}(\cdot, \cdot)$ as

$$\mathcal{D}(v_i, v_j) = \min_{\Pi_{i,j}} \sum_{e_{ij} \in \Pi_{i,j}} w(e_{ij}), \tag{8.1}$$

where $w(e_{ij})$ is the distance weight of edge e_{ij}, $\Pi_{i,j}$ is a set of edges representing a path between v_i and v_j, and we define $\mathcal{D}(v_i, v_i) = 0$. The optimal path connecting the pair of nodes is called the **shortest path**, and thus the distance is defined as the length of the shortest path between the pair of nodes along the edges of the graph. If no path connects v_i and v_j (the graph is disconnected), then we define the corresponding distance as $\mathcal{D}(v_i, v_j) = \infty$. When the weights are positive, the distance defined in this way establishes a formal *metric* between nodes on the graph, since the distance is nonnegative, symmetric, discernible and satisfies the triangle inequality (see Sect. 4.4).

A **metric** can be defined for any graph with positive edge weights by defining the distance between any pair of nodes by the length of the shortest path connecting them.

There are many fast algorithms available to compute the distance between two nodes. The most common algorithm for computing distance is Dijkstra's algorithm [108] which may be applied to any connected graph with nonnegative weights.

8.1.2 Node Centrality

Armed with the notion of distance between two nodes given by (8.1), we may examine the importance of a particular node by considering the distance between the node and the rest of the network. Measures of node importance are called **node centrality** measures due to the use of distances to determine how "central" a node is within a network (for a more extensive review of node centrality, see [247]). The most natural method for using distance to measure the centrality of a particular node is to examine the distance from the node to all other nodes in the graph. This measure of node centrality is known as the **total distance** of a node v_i, which is defined as

$$\text{TD}(v_i) = \sum_{v_j} \mathcal{D}(v_i, v_j). \tag{8.2}$$

The total distance is proportional to the **closeness** measure of a node, which is defined as the average distance from the node to all other nodes in the graph, i.e.,

$$\text{Closeness}(v_i) = \frac{1}{n-1}\,\text{TD}(v_i), \tag{8.3}$$

where n is the number of nodes in the graph, $n = |\mathcal{V}|$. Total distance for a single node may be computed efficiently by Dijkstra's algorithm.

The total distance and closeness both measure node centrality by examining the distance between a single node and the entire network. Therefore, each quantity provides an aggregated measure of node centrality. Instead of examining an aggregated measure of node centrality, we could adopt another view of node centrality by considering only the worst-case distance between a node and the remaining network. A measurement of worst-case node centrality is provided by the node **eccentricity**, which measures the maximum distance between the node and any other node in the network. Eccentricity is defined for node v_i as

$$\text{Eccentricity}(v_i) = \max_{v_j}\ \mathcal{D}(v_i, v_j). \tag{8.4}$$

A different approach to measuring node centrality is to examine the importance of a node as a link between other pairs of nodes. This **betweenness** has been characterized numerically as

$$\text{Betweenness}(v_i) = \sum_{\substack{v_j, v_k \\ v_i \neq v_j \neq v_k}} \frac{\sigma_{v_j, v_k}(v_i)}{\sigma_{v_j, v_k}}, \tag{8.5}$$

where $\sigma_{v_j, v_k}(v_i)$ indicates the number of shortest paths between v_j and v_k that pass through v_i, and σ_{v_j, v_k} indicates the total number of shortest paths joining v_j and v_k. Therefore, the betweenness measure represents the fraction of optimal paths between every pair of nodes that cross through v_i. Calculation of node betweenness is more expensive than the previous measures due to the fact that it is necessary

to know the optimal paths between all pairs of nodes. The classic algorithm for computing all-pairs shortest paths is the Floyd–Warshall algorithm [87], although Johnson's algorithm or repeated applications of Dijkstra's algorithm are considered more efficient for sparse graphs [87]. Due to the usefulness of the betweenness measure in practice, specialty algorithms to compute betweenness have been developed that are faster than computation of the all-pairs shortest paths (the most efficient specialty algorithm is by Brandes [57]). However, even these specialty algorithms for computing betweenness are too expensive for large graphs, causing continued work on fast algorithms to approximate betweenness [16].

Not every measure of node centrality is based on shortest-path distance. Another common measure of node centrality is simply the number of neighbors of the node (the node degree). Degree may be a useful measure of node centrality for networks in which there is a large range of different node degrees, but may be less useful to describe the importance of nodes in networks for which the degree distribution has low variance.

8.1.3 Distance-Based Properties of a Graph

Having explored methods for measuring the centrality of nodes from distance calculations, we may use these methods to define a series of distance-based measures to describe the "connectedness" of the entire graph. The first of these global measures is the most successful measure used in chemical graph theory, and was defined by Wiener. Although many topological indices have since been proposed in the chemical graph theory literature, the measure used by Wiener has had the most success and influence. This measure is known as the **Wiener index** or **Wiener number** which is defined for a graph $\mathcal{G}$ as

$$W(\mathcal{G}) = \frac{1}{2} \sum_{v_i} \sum_{v_j} \mathcal{D}(v_i, v_j). \tag{8.6}$$

The Wiener index represents the sum of the shortest path lengths between all pairs of nodes in the graph. Therefore, a graph with a small Wiener index is more well-connected than a graph (with the same number of nodes) having a large Wiener index. In the network theory literature, it is more common to use the **average path length** which normalizes the distances comprising the Wiener index by the number of node pairs. The average path length is defined as

$$\text{AveragePathLength}(\mathcal{G}) = \frac{2W(\mathcal{G})}{n^2 - n} = \frac{1}{n^2 - n} \sum_{v_i} \sum_{v_j} \mathcal{D}(v_i, v_j). \tag{8.7}$$

Calculation of the Wiener index and average path length requires knowledge of the shortest paths between all vertex pairs in a network. Consequently, the Floyd–Warshall algorithm or the Johnson algorithms [87] are the most common methods for calculating the Wiener index and average path length.

The average path length and the Wiener index are measures of the connection strength between all pairs of nodes in the graph. However, both of these measures define the connection strength between nodes by the length of the *optimal* path. Since these measures are based purely on the length of a single optimal path, they may not reflect a more global measure of connectedness in the graph defined by multiple paths. Therefore, a different view of node connectivity is to define the strength of a connection between two nodes by the number of parallel paths. One way of measuring the strength of parallel connections is to use the resistance distance [235] between two nodes when the graph is viewed as an electrical circuit (see Chap. 3 for more discussion of the effective resistance and circuit analogy). Specifically, the resistance distance has been used to measure the strength of parallel paths connecting all pairs of nodes by defining the **Kirchhoff index** [43, 109]

$$\mathrm{KI}(\mathcal{G}) = \frac{1}{2} \sum_{v_i} \sum_{v_j} R_{\mathrm{eff}}(v_i, v_j) = n \,\mathrm{trace}\{\mathbf{L}^{\dagger}\}, \tag{8.8}$$

where $\mathbf{L}^{\dagger}$ indicates the pseudoinverse of the Laplacian matrix and $R_{\mathrm{eff}}(v_i, v_j)$ is the effective resistance between nodes v_i and v_j, as defined in Sect. 4.4. The Kirchhoff index of any graph is always smaller than the Wiener index, since $\mathcal{D}(v_i, v_j) \geq R_{\mathrm{eff}}(v_i, v_j)$ implies that $W(\mathcal{G}) \geq \mathrm{KI}(\mathcal{G})$. This inequality becomes an equality when there exists only a single path between all pairs of nodes, i.e., the graph is a tree.

When computing the Wiener index or Kirchhoff index for a weighted graph, care must be taken in the interpretation of the weights. Weights are incorporated in the definition of the distance measure (8.1) required by the Wiener index, and weights are also incorporated in the definition of the Laplacian matrix (2.96) required by the Kirchhoff index. However, in Chap. 2 we saw that the roles of the prescribed weights are *not the same* in these two cases. Given a weighted graph, one must take care to choose, depending on the problem and on the origin of the weights, whether the weights are "distance weights" or "affinity weights". Chapter 2 contains a discussion of these two interpretations of weights, and there it is shown that distance weights correspond to the primal metric tensor and affinity weights to the dual metric tensor. If the measures of distance are to be equivalent when quantified using the distance operator and using the Laplacian operator, distance weights must be used in the definition of the distance measure, and affinity weights in the definition of the Laplacian.[1] Specifically, for the two measures to operate using the same underlying metric on a single graph, i.e., for the two measures to be compatible, the weights used in the Laplacian matrix must be the *reciprocal* of the prescribed edge weights used in the distance measure. For example, in the case of a tree the two indices should be equal. However, improper interpretation of the weights will cause the two calculated indices to be unequal (i.e., using distance weights for both measures or

[1] In the circuit theory analogy, for the two indices to be compatible the prescribed edge weights of a weighted graph are interpreted as *resistances* when measuring the Wiener index, and the same prescribed edge weights on the weighted graph are interpreted as *conductances* when measuring the Kirchhoff index.

affinity weights for both measures will not yield the same value). Therefore, the comparison of the Wiener index and the Kirchhoff index provides a good example of the importance of distinguishing these two interpretations of the prescribed edge weights.

The **quasi-Wiener index** was also introduced [109, 279] to measure the strength of parallel paths connecting all pairs of nodes by calculating the eigenvalues of the Laplacian matrix. However, Gutman and Mohar proved that the quasi-Wiener index and the Kirchhoff index are equal for all graphs [182] and therefore we do not discuss the quasi-Wiener index any further.

Although the authors are not aware of any attempts to use topological indices defined on the *edge Laplacian* to measure other aspects of the graph (or the chemical properties of the molecule), the measures above suggest an easy extension to global edge–face (edge–cycle) relationships. Specifically, a higher-order Kirchhoff index could be defined as

$$\text{Higher-OrderKI}(\mathcal{G}) = m \,\text{trace}\{\mathbf{L}_1^{\dagger}\}, \tag{8.9}$$

where $m = |\mathcal{E}|$ and $\mathbf{L}_1$ is the edge Laplacian (see Chap. 2). If the graph cycle set constitutes a basis (see Chap. 4), then the edge Laplacian matrix has full rank and the pseudoinverse is replaced by the true inverse of the edge Laplacian matrix. Recall from Chap. 4 that a set of $|\mathcal{E}| - |\mathcal{V}| + 1$ independent cycles will form a basis. This measure of the higher-order Kirchhoff index does not retain its original interpretation in terms of the effective resistance, since the effective resistance between two edges does not have a conventional definition. Despite losing this interpretation, this measure on the edge Laplacian matrix provides a value that indicates the relative "connectedness" of the edges via their incident nodes and cycles.

The Wiener index, average path length and Kirchhoff index all provide an aggregate measure of distance between all nodes in the graph. As before, we may use the definition of node eccentricity to define a measure of worst-case distance between all nodes. Specifically, the definition of node eccentricity allows for the definition of the **graph radius** and **graph diameter**

$$\text{Radius}(\mathcal{G}) = \min_{v_i} \text{Eccentricity}(v_i), \tag{8.10a}$$

$$\text{Diameter}(\mathcal{G}) = \max_{v_i} \text{Eccentricity}(v_i). \tag{8.10b}$$

A node for which $\text{Eccentricity}(v_i) = \text{Radius}(\mathcal{G})$ is called a graph **center** and a node for which $\text{Eccentricity}(v_i) = \text{Diameter}(\mathcal{G})$ is called a **peripheral node**.

Figure 8.1 gives an example of these distance-based measures used to describe a tree and a small lattice. The distance-based measures considered in this section have all been applied to measure some aspect of connectedness in a graph. In the next section we consider measures of graph *separability*.

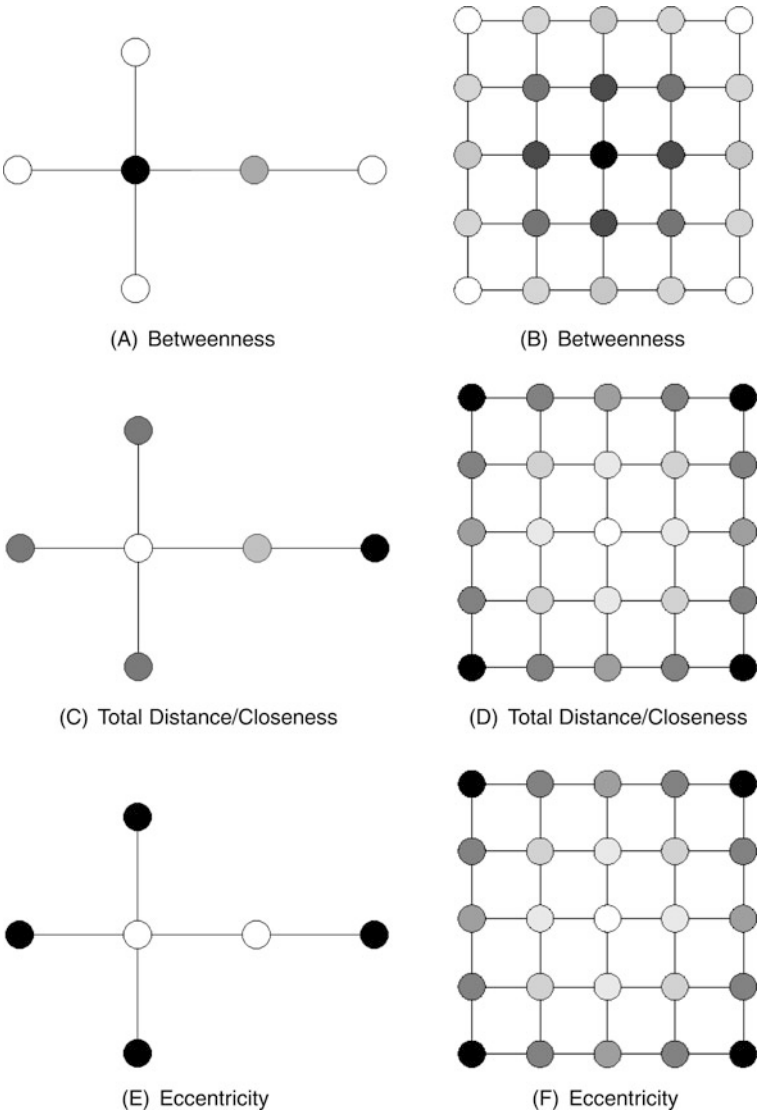


Fig. 8.1 Distance measures on two example graphs, displaying betweenness, TD and eccentricity for each node. Greater values of these quantities are represented by *darker shading*. Each example uses normalized values for the node quantities and unit weights for all edges. (*Left*) Measures for tree example: Radius = 2, Diameter = 3, Wiener Index = 28, Average Path Length = 1.8667, Kirchhoff Index = 28, Higher-Order Kirchhoff Index = 0. (*Right*) Measures for lattice example: Radius = 4, Diameter = 8, Wiener Index = 1000, Average Path Length = 3.333, Kirchhoff Index = 338.03, Higher-Order Kirchhoff Index = 814.42. From the node eccentricity examples we can determine which nodes are centers and which nodes are peripheral. *Tree graph*: all nodes with one neighbor are peripheral and the other two nodes are centers. *Lattice graph*: the corner nodes are peripheral and the middle node is the only center

8.2 Measures of Graph Separability

Similar to the distance-based measures of graph connectedness presented above, measures of graph separability are built upon metric properties of the graph. We first consider measures based on the volume of sets in a graph, then consider an example of separability based on distances.

8.2.1 Clustering Measures

Clustering measures use partitions of a graph to determine the separability of the graph. One of the most ancient measures of separability of a space arises from the **isoperimetric problem**, which seeks the shape with largest area/volume from the set of all shapes with the same perimeter/surface area. In the continuous, Euclidean $\mathbb{R}^3$ domain, the solution of this problem is known to be the sphere, and in $\mathbb{R}^2$ the circle. In a finite space (e.g., the surface of a closed object), one may define the **isoperimetric ratio** of an N-dimensional closed Riemannian manifold, $\mathcal{M}$, as [74]

$$h(\mathcal{M}) = \frac{\partial \mathcal{S}}{\min(\text{Vol}(\mathcal{S}), \text{Vol}(\bar{\mathcal{S}}))}, \tag{8.11}$$

where $\mathcal{S}$ represents an N-dimensional submanifold, $\bar{\mathcal{S}}$ represents its complement and $\partial\mathcal{S}$ represents the boundary length (surface area) of $\mathcal{S}$. Instead of fixing a perimeter, $\partial\mathcal{S}$, and seeking the submanifold $\mathcal{S}$ with greatest volume, we may generalize the isoperimetric problem to seek the node set $\mathcal{S} \subset \mathcal{V}$ that minimizes the isoperimetric ratio. Such a solution and its complement are called the **isoperimetric sets** of the domain. The value of the minimum isoperimetric ratio gives a notion of separability of the space. For example, if the domain were disconnected into two pieces, then the isoperimetric sets would consist of each piece individually and the isoperimetric ratio would be zero. Similarly, the solution to the isoperimetric problem on the surface of a "dumbbell" is also the two balls of the dumbbell with a small neck separating them [74]. Since the boundary of $\mathcal{S}$ would be measured on the neck, and the surface area on the two balls, the isoperimetric ratio of a dumbbell is small, meaning that a dumbbell is *nearly disconnected*. As a measure of separability for a domain, we may also consider the isoperimetric ratio of a graph. To do so, we must define the analogous concepts used above for a graph. Specifically, we let the set $\mathcal{S}$ refer to a set of nodes such that $\mathcal{S} \subset \mathcal{V}$, $\mathcal{S} \neq \emptyset$, $|\mathcal{S}| \leq \frac{1}{2}|\mathcal{V}|$ with boundary defined as the sum of the weight of edges spanning the complementary subsets $\mathcal{S}$ and $\bar{\mathcal{S}}$. Two definitions of the **volume** of a set of nodes in a graph are defined as [81, 285]

$$\text{Vol}_1(\mathcal{S}) = |\mathcal{S}|, \tag{8.12}$$

or

$$\text{Vol}_2(\mathcal{S}) = \sum_{v_i \in \mathcal{S}} d_i, \tag{8.13}$$

where d_i represents the degree of node v_i. These two definitions of volume have led to two different definitions of the isoperimetric ratio of a graph.

We may write the isoperimetric ratio of the set $\mathcal{S}$ using an indicator vector, i.e., the vector $\mathbf{x}$ such that $x_i = 1$ if $v_i \in \mathcal{S}$ and $x_i = 0$ otherwise. Using the indicator vector allows us to write the isoperimetric ratio as either

$$h_1(\mathcal{S}) = \frac{\partial \mathcal{S}}{\text{Vol}_1(\mathcal{S})} = \frac{\mathbf{x}^\mathsf{T}\mathbf{L}\mathbf{x}}{\mathbf{x}^\mathsf{T}\mathbf{x}} \tag{8.14}$$

or

$$h_2(\mathcal{S}) = \frac{\partial \mathcal{S}}{\text{Vol}_2(\mathcal{S})} = \frac{\mathbf{x}^\mathsf{T}\mathbf{L}\mathbf{x}}{\mathbf{x}^\mathsf{T}\mathbf{D}\mathbf{x}}. \tag{8.15}$$

As in the continuous case, the minimum isoperimetric ratio over all possible sets $\mathcal{S}$ gives a measure of the separability of the graph. This minimum of $h(\mathcal{S})$ over all possible $\mathcal{S}$ is called the **isoperimetric number**, **isoperimetric constant** or **Cheeger constant**. We note that there is some disagreement in the literature about these definitions, since all of these terms have been applied by various authors to either $h_1(\mathcal{G})$ or $h_2(\mathcal{G})$. We will use the term *isoperimetric constant*, which we denote as $h_1(\mathcal{G})$ or $h_2(\mathcal{G})$, depending on the definition of volume. The quantity $h_1(\mathcal{G})$ for a graph is additionally known as the **edge expansion** and the quantity $h_2(\mathcal{G})$ is known as the **graph conductance**.

The isoperimetric constant of a graph appears frequently in the literature. For example, the isoperimetric constant has been used to characterize **expander graphs** [6, 7]. Additionally, the **graph partitioning problem** is often formulated explicitly with the goal of finding the isoperimetric sets [169]. Unfortunately, calculation of the minimum isoperimetric ratio of an arbitrary graph is NP-Hard [285, 286]. Therefore, one approach to estimating the isoperimetric number of a graph is to apply several different graph partitioning algorithms and use the smallest isoperimetric ratio of these partitions as an estimate of the isoperimetric constant. Figure 8.2 shows examples of isoperimetric sets on two graphs, one of which exhibits a natural clustering and the other which does not.

A different approach to estimating the minimum isoperimetric constant of a graph is to use known bounds for the constant. The standard method for bounding the isoperimetric constant is through use of the **Fiedler value**, which is defined as the smallest nonzero eigenvalue of the Laplacian matrix. Fiedler observed that this eigenvalue was a good predictor of graph separability. Based on this observation, Fiedler named the smallest nonzero eigenvalue of the Laplacian matrix the **algebraic connectivity** [133] (although it is now known as the Fiedler value). The algebraic connectivity is a meaningful measurement of the graph separability, but it may also be used to bound the minimum isoperimetric constant $h_1(\mathcal{G})$ from both above and below. Specifically, the following expression combines **Cheeger's inequality** (upper bound) and **Buser's inequality** (lower bound) [74, 81] to provide

$$\sqrt{2d_{\max}\lambda_2} \geq h_1(\mathcal{G}) \geq \tfrac{1}{2}\lambda_2, \tag{8.16}$$

where $d_{\max}$ is the maximum degree in the graph and λ_2 is the Fiedler value. Note that the these inequalities also hold for graphs with any set of positive edge weights. The right side of this equality may be seen easily since a real-valued relaxation of $\mathbf{x}$ causes (8.14) to be an expression for the Rayleigh quotient of $\mathbf{L}$, which is minimized by λ_2 for all solutions of $\mathbf{x}$ orthogonal to the constant vector. Similarly, for $h_2(\mathcal{G})$ we have

$$\sqrt{2\lambda_2^*} > h_2(\mathcal{G}) \geq \tfrac{1}{2}\lambda_2^*, \tag{8.17}$$

where λ_2^* indicates the smallest nonzero eigenvalue of the *normalized* Laplacian matrix. The appearance of the normalized Laplacian matrix is not surprising if we view the two definitions of volume in (8.12) and (8.13) as different definitions of *node weights*. Defining a node weight as unity (8.12) or defining a node weight as the node degree (8.13) lead naturally to the standard unweighted Laplacian matrix or the normalized Laplacian matrix, respectively (see Chap. 2 and [111, 112] for more details).

In addition to the node Fiedler value, we may define a *higher-order* Fiedler value defined on edges via the edge Laplacian. The higher-order Fiedler value is therefore defined as the smallest eigenvalue of the edge Laplacian. If the complex is simply connected (i.e., the number of faces included in the complex is greater than zero and equals $m - n + 1$), then this higher-order Fiedler value is positive. However, if there are fewer than $m - n + 1$ cycles in the complex, then the higher-order Fiedler value would correspond to the first nonzero eigenvalue (in accordance with the node case).

Another approach to measuring separability of a graph is the **clustering coefficient**, which looks locally for evidence of separability. The clustering coefficient is computed for a node by counting the number of pairs of neighbors of a node that are also neighbors of each other. This concept is motivated by social networks in which it has been observed that if one individual (node) knows two other people (has edges connecting them) then these people are likely to know each other (be connected via an edge). Since this set of connections close a triangle in the graph, this concept is sometimes known as **triadic closure**. Therefore, a pair of neighbors is considered *closed* if these neighbors are connected. With these definitions, the clustering coefficient at node v_i is computed as

$$\mathrm{CC}(v_i) = \frac{\text{Number of closed pairs of neighbors of } v_i}{\text{Total number of pairs of neighbors of } v_i}. \tag{8.18}$$

Since $\mathrm{CC}(v_i)$ is undefined if $d_i < 2$, we adopt the convention that $\mathrm{CC}(v_i) = 0$ for these nodes. The clustering coefficient for the entire graph is given by

$$\mathrm{CC}(\mathcal{G}) = \frac{1}{n}\sum_i \mathrm{CC}(v_i). \tag{8.19}$$

Therefore, this value takes a maximum at unity when $\mathcal{G}$ is fully connected and a minimum of zero when $\mathcal{G}$ represents a tree. Unfortunately, this definition of the clustering coefficient does not account for the graph weights. Several possibilities

for extending this definition to include graph weights were proposed in [298] and an extension of this measure beyond triangles was given in [230]. Examples of the clustering coefficient and its comparison to the isoperimetric sets are provided in Fig. 8.2.

8.2.2 Small-World Graphs

The intuitive concepts of distance-based *connectedness* and *separability* seem to be opposite, i.e., that a separable network is not well-connected and *vice versa*. However, it is known that social networks contain clear groups of tightly coupled people but, despite this grouping, that the entire network is connected by short optimal paths. Watts and Strogatz explained this phenomenon with a model known as a **small-world network** [362, 396]. A small-world network was defined as any graph that has both a small average path length as defined in (8.7) and a large clustering coefficient as defined in (8.19). The name "small-world" network is derived from the social experiments of Milgram [283] who found that it was possible to send messages through a large, locally clustered, social network with approximately six steps. This result brought the concept of *six degrees of separation* to popular culture. Since the introduction of small-world networks, these models have been widely used to produce an understanding of many phenomena in physics, sociology and biology [397].

A key observation of Watts and Strogatz was that the average path length of a *random graph* is small when the graph is connected. Random graphs were initially (and comprehensively) studied by Erdős and Rényi [124, 125]. A random graph is defined as a graph which starts as a set of disconnected nodes to which k edges are progressively added. The added edges are chosen randomly from the set of all possible edges (node pairs), excluding multiple and self connections. Watts and Strogatz used the results on average path length to show that any locally connected graph (i.e., a graph with a large clustering coefficient) could be converted into a small world graph by randomly rewiring a small number edges. This randomization of a small number of edges has the effect of drastically lowering the average path length. This technique for producing a graph with small average path length has been exploited in the context of improving the speed of graph-based computer vision algorithms [167].

Following the publication of the initial work on small-world networks, other types of small-world networks have also been studied. A motivation for studying additional classes of small-world networks was the fact that the initial Watts–Strogatz model did not resemble many real networks in the sense that most nodes in the Watts–Strogatz model have the same degree. In fact, the **degree distribution** (the histogram of node degrees) of many real networks follows a power law [20] (e.g., the world wide web [3]). A network with a degree distribution following a power law is called a **scale-free network**. Recent interest in scale-free networks was generated by the work of Barabási and Albert [20], who suggested a mechanism for

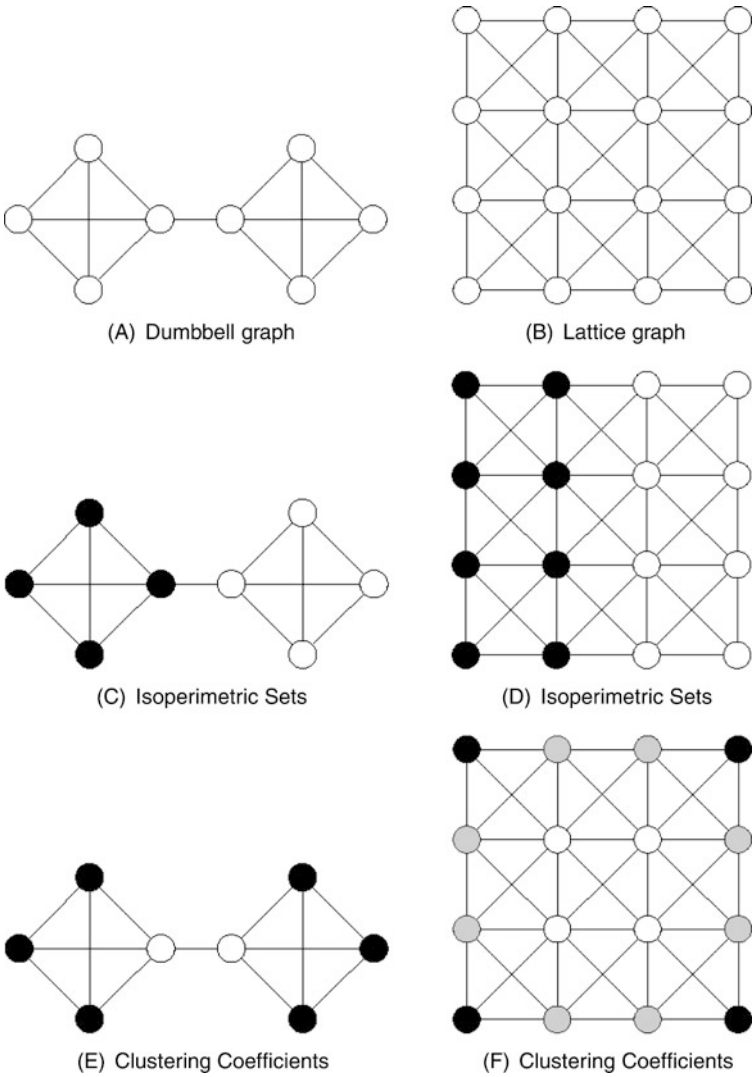


Fig. 8.2 Two examples of the clustering measures. The dumbbell graph on the *left* clusters well while the 8-connected lattice on the *right* does not. Node membership in the isoperimetric sets are indicated by coloring the nodes as *white* or *black*. Clustering coefficients are displayed for each node via *shading*, where *darker shading* represents nodes with a higher clustering coefficient. A node with a higher clustering coefficient may be interpreted as indicating that the node is more "interior". Dumbbell global clustering measures: $\text{Vol}_1 = 8$, $\text{Vol}_2 = 26$, $h_1 = 0.25$, $h_2 = 0.0769$, Fiedler Value $= 0.3542$, Graph Clustering Coefficient $= 0.8750$, Higher-Order Fiedler Value $= 0.3542$. Lattice global clustering measures: $\text{Vol}_1 = 16$, $\text{Vol}_2 = 84$, $h_1 = 1.25$, $h_2 = 0.2381$, Fiedler Value $= 1.4364$, Graph Clustering Coefficient $= 0.6571$, Higher-order Fiedler Value $= 1.0$. These measures tell us that, compared to the dumbbell, the lattice graph is larger (greater volume), harder to separate (larger isoperimetric constants and Fiedler value), and fewer of the neighbors of each node are connected (smaller clustering coefficient)

producing a scale-free network. This mechanism is known as **preferential attachment** and roughly states that as edges are added to a graph, the nodes with larger degree have an increased probability of being linked. This process therefore produces some nodes with very large degree (called "hubs") and the remaining degree distribution follows a power law [20]. An example of this type of process is in the author citation network, in which a paper with many citations is more likely to continue to be cited in the future. Although scale-free networks are commonly used as examples of small-world networks, the clustering coefficient is not always large (although the diameter is small), meaning that scale-free networks are not necessarily small-world graphs.

The measures of graph connectedness and graph separability considered thus far are measures that naturally pertain to 1-complexes or graphs. We now consider global topological measures for general p-complexes, and afterwards consider geometric measures defined specifically for surfaces or 2-complexes.

8.3 Topological Measures

The distance-based measures that we have studied so far are often called "topological indices" in the chemical graph theory literature. Although these methods do measure aspects of the network topology, they do not measure the usual topological invariants such as the Euler characteristic, genus, Betti numbers, torsion coefficients and orientability. In this section, we will describe how to calculate these invariants for a cell complex. Here we assume that we are measuring the topological properties of a p-complex. Recall from Chap. 2 that a p-complex is defined by sets of p-dimensional cells, $\mathcal{S}_p$. A standard graph is therefore a 1-complex, which contains sets of nodes and edges only. The subject of computational topology on a complex is treated extensively in the literature [119, 254, 425].

We begin our treatment of topological measures with a discussion of the Betti numbers for a complex. The pth Betti number of an n-complex is defined as the rank of the pth homology group [254]. Informally, the Betti number may be viewed as the number of cuts that may be made without dividing a surface into two parts. Therefore, when $p = 0$, the Betti number represents the number of connected components, when $p = 1$, the Betti number represents twice the number of handles, and for $p = 2$, the Betti number represents the number of voids. Computation of the pth Betti number is possible from the incidence matrices via the formula [254]

$$\text{Betti}_p = |\mathcal{S}_p| - \text{Rank}(\mathbf{N}_{p+1}) - \text{Rank}(\mathbf{N}_p), \tag{8.20}$$

where we may recall that $\mathbf{N}_p$ represents the pth incidence matrix of the complex. To make this definition hold for all p on a p-complex, we define $\text{Rank}(\mathbf{N}_p) = 0$ for $p \leq 0$ or for $p \geq n$. As an example, we may give the calculation of the Betti numbers for $p = 0$ and $p = 1$ using our conventional notation as

$$\text{Betti}_0 = |\mathcal{V}| - \text{Rank}(\mathbf{A}) - 0, \tag{8.21}$$

$$\text{Betti}_1 = |\mathcal{E}| - \text{Rank}(\mathbf{B}) - \text{Rank}(\mathbf{A}). \tag{8.22}$$

Recall that Rank($\mathbf{A}$) equals $|\mathcal{V}| - c$ where c represents the number of connected components (see Chap. 2). Therefore, $\text{Betti}_0 = c$. Similarly, when a complex is closed and simply connected, then we saw in Chap. 2 that Rank($\mathbf{B}$) $= |\mathcal{E}| - |\mathcal{V}| + 1$. Therefore, when $c = 1$ and the complex is closed and simply connected, $\text{Betti}_1 = 0$.

A simpler method for calculating the Betti numbers was given by Friedman [143] who noted that, by the definition of the pth-order Laplacian, it was possible to calculate the Betti numbers as

$$\text{Betti}_p = \text{Dimension}(\text{Nullspace}\{\mathbf{L}_p\}). \tag{8.23}$$

This approach to calculating the Betti numbers is often more straightforward since it may be computed by counting the number of zero eigenvalues of $\mathbf{L}_p$. This second expression for the Betti numbers in (8.23) also recovers the fact that $\text{Betti}_0 = c$, since it is well-known [34] that Rank(Nullspace$\{\mathbf{L}_0\}$) $= c$.

Multiplying the general expression for the Betti numbers in terms of the ranks of incidence matrices given in (8.20) on both sides by $(-1)^p$, results in the Euler–Poincaré theorem which states that

$$\sum_{i=0}^{p} (-1)^p |\mathcal{S}_i| = \sum_{i=0}^{p} (-1)^p \, \text{Betti}_i. \tag{8.24}$$

The value of this sum is known as the **Euler characteristic** of a p-complex, $\chi(\mathcal{G})$, i.e.,

$$\chi(\mathcal{G}) = \sum_{i=0}^{p} (-1)^p |\mathcal{S}_i| = \sum_{i=0}^{p} (-1)^p \text{Betti}_i. \tag{8.25}$$

When $p = 2$, this equation gives the usual formula for a surface $\chi(\mathcal{G}) = |\mathcal{F}| - |\mathcal{E}| + |\mathcal{V}|$. The Euler characteristic is one of the central invariants in topology. For a connected complex, the Euler characteristic may be considered as the number of linearly independent cells which are possible, but not necessarily present, in the complex (plus one). For example, we have seen that the number of linearly independent cycles in a graph are equal to $|\mathcal{E}| - |\mathcal{V}| + 1$. Therefore, if all of these cycles are included in our set of faces, then $|\mathcal{F}| = |\mathcal{E}| - |\mathcal{V}| + 1$ and the Euler characteristic equals one. If the "exterior face" is additionally included in the set of faces (see Chap. 2) then the Euler characteristic equals two, which is the classical result for a simply-connected closed surface.[2] Two examples are given in Fig. 8.3 showing a cube with Euler characteristic two and an annulus with Euler characteristic zero.

[2]The exterior face is a device to enable a finite graph to be defined such that it has no boundary and is therefore closed. This imparts a global topology on the graph—that of a sphere or a sphere with handles—which may be interpreted as a finite graph including a face "at infinity", in analogy to projections of the sphere into the plane (e.g., by stereographic projection) in which the coordinate at the pole of the sphere opposite the origin is mapped to the point at infinity on the flat plane.

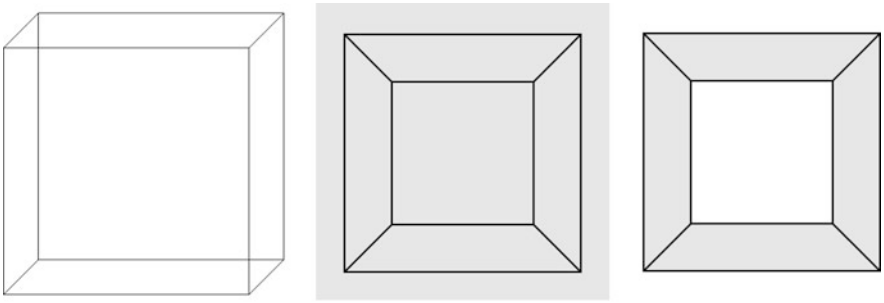


Fig. 8.3 A cube, the flattened cube (with exterior face), and an annulus. The Euler characteristic is $\chi = 6 - 12 + 8 = 2$ for the cube (flattened and not flattened), and $\chi = 4 - 12 + 8 = 0$ for the annulus. Faces included in the face set are shaded *gray* (including the exterior face for the flattened cube)

Closely related to the Euler number of a closed, orientable 2-complex is its **genus**, which may be defined in terms of the Euler characteristic via the relationship

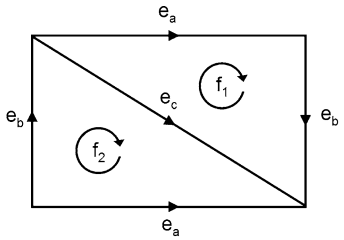
$$\chi(\mathcal{G}) = 2 - 2\,\text{Genus}(\mathcal{G}) \tag{8.26}$$

for a closed surface. The genus is often thought of as the number of handles on the surface. Therefore, the sphere has genus equal to zero, whereas the torus (a sphere with a handle) has a genus equal to one. The genus of a 1-complex (graph) has been defined as the minimum genus of the surface on which the graph may be embedded such that is has no edge crossings [188]. Therefore, a planar graph is considered to have genus zero.

The property of surface **orientability** describes whether or not it is possible to describe the surface as the boundary of some object. A classical example of a non-orientable structure is the Möbius strip. One test for surface orientability on a finite complex is to examine the orders of the torsion subgroups of the homology group of the complex, which are known as the **torsion coefficients** of the complex [254, 272]. These coefficients may be determined computationally by calculating the invariant factors of the incidence matrix (i.e., the diagonal of the incidence matrix after placing it in Smith Normal Form). Specifically, the kth torsion coefficients are defined as the set of invariant factors of the kth incidence matrix greater than unity [254]. Therefore, if the complex is orientable, then the complex is torsion-free and there are no invariant factors of the kth incidence matrix greater than unity [272]. For example, Fig. 8.4 shows a triangulated Klein bottle with its corresponding face–edge incidence matrix and the incidence matrix in Smith Normal Form.

8.4 Geometric Measures

In addition to the properties considered above, we may also compute classical *geometric* quantities to describe a complex. Here we follow the computer graphics



(A) Klein bottle cell complex

$$\begin{array}{c} \\ e_a \\ e_b \\ e_c \end{array} \begin{array}{c} \begin{array}{cc} f_1 & f_2 \end{array} \\ \begin{bmatrix} 1 & -1 \\ 1 & 1 \\ -1 & 1 \end{bmatrix} \end{array}$$

(B) Face-edge incidence matrix

$$\begin{array}{c} \\ e_a \\ e_b \\ e_c \end{array} \begin{array}{c} \begin{array}{cc} f_1 & f_2 \end{array} \\ \begin{bmatrix} 1 & 0 \\ 0 & 2 \\ 0 & 0 \end{bmatrix} \end{array}$$

(C) Smith Normal Form

Fig. 8.4 Torsion coefficients may be extracted from the face–edge incidence matrix by placing the matrix in Smith Normal Form. In this example, we give the face–edge incidence matrix of a cell decomposition of the (flattened) Klein bottle [196] and show that the Klein bottle has a torsion coefficient equal to two since the only invariant factor greater than unity lies at face f_2

literature which defines measures of curvature for a 2-complex which is embedded in $\mathbb{R}^3$.

Geometric measures defined on simplicial complexes (i.e., triangular meshes representing surface for 2-complexes) depend on the metrics ascribed to the complexes. In the case of discrete surfaces the metric is typically derived from the embedding of the two-dimensional surface in a three-dimensional metric space, although the notion of distance may be produced from any process. Therefore, the distances along the surface are the natural Euclidean distances inherited from the ambient space of the embedding.

Here we will consider two forms of curvature defined on surfaces: Gaussian and mean curvature. **Gaussian curvature** is an example of an *intrinsic* geometric property of the surface in that Gaussian curvature is invariant under isometric transformations of the surface. That is, if the surface is deformed in a way that does not affect the distance between any pair of vertices on the surface, then the Gaussian curvature is also unaffected. For this reason, the Gaussian curvature is called intrinsic and depends only on the metric of the surface. **Mean curvature**, however, is an example of an *extrinsic* geometric property that can change under isometric transformations. For instance, a punctured sphere and a flat disk in the plane are topologically equivalent, and therefore a smooth homeomorphism exists between them, but a sphere has constant positive Gaussian curvature and a disk has zero Gaussian curvature—it is clear that the deformation between the two configurations cannot take place without geometric distortion. However, the cylindrical tube is an example of a surface with zero Gaussian curvature and non-zero mean curvature, so if the tube is cut straight down its side (imagine a piece of paper rolled so that the two short ends meet) it can be unrolled and flattened into the plane without stretching or intrinsic geometric distortion.

8.4.1 Discrete Gaussian Curvature

Gaussian curvature is typically defined in terms of the principle curvatures [110]. However, an alternate definition of Gaussian curvature is provided by a special case

of the Gauss–Bonnet theorem that applies equally well to the continuous or discrete cases. This alternate definition of Gaussian curvature is based on the idea that, in the plane, at a given vertex the sum of angles between adjacent edges connecting the vertex to its neighbors always sum to 2π. For curved surfaces, this same sum can be either greater than or less than 2π, and this *angle deficiency* is the basis of Gaussian curvature.

The Gauss–Bonnet theorem establishes a deep result in differential geometry that links the intrinsic geometry of a manifold $\mathcal{M}$ to its topology. For a given metric, the integral of the Gaussian curvature K over the manifold (also known as the **total curvature**), plus a boundary term consisting of the integral of the geodesic curvature, k_g, along the manifold boundary, is related to the Euler characteristic of the manifold as

$$\iint_{\mathcal{M}} K \, \mathrm{d}V + \int_{\partial \mathcal{M}} k_g \, \mathrm{d}s = 2\pi \, \chi(\mathcal{M}). \tag{8.27}$$

Therefore, for any topological sphere, the total curvature is always 4π, whereas the total curvature for a 1-torus is always 0.

This general theorem can be applied locally to provide an integral definition for Gaussian curvature on two-dimensional surfaces that holds for discrete spaces. If we consider the total curvature of the dual cell surrounding each node of a graph, then for a particular node v_i at its corresponding dual cell $\mathcal{C}$ the Gauss–Bonnet theorem reduces to

$$\iint_{\mathcal{C}} K \, \mathrm{d}V - \int_{\partial \mathcal{C}} k_g \, \mathrm{d}s = 2\pi \tag{8.28}$$

since, by definition, $\chi(\mathcal{C}) = 1$ for the cell. For the case in which the dual cell is conveniently given by the Voronoï cell at each vertex, the integral of the geodesic curvature is zero at the piecewise linear edges of the cell boundary plus the angles at the corners of its boundary. Furthermore, these angles are equivalent to the interior angles θ at node v_i formed between each pair of edges incident to v_i in this special case of a Voronoï cell [280, 366]. As a result, the Gaussian curvature $K(v_i)$ at node v_i may be neatly defined as

$$K(v_i) = \frac{2\pi - \sum_{j \, \forall e_{ij}} \theta_j}{A_{v_i}} \tag{8.29}$$

where θ_j is given by $\theta_j \equiv \angle\{v_j, v_i, v_{(j+1) \bmod B}\}$ if B represents the number of neighboring nodes, and A_{v_i} represents the area of the Voronoï cell.

8.4.2 Discrete Mean Curvature

The definition of Gaussian curvature at a vertex in (8.29), which phrases the integral of the curvature over a small patch as the deviation from 2π of the internal

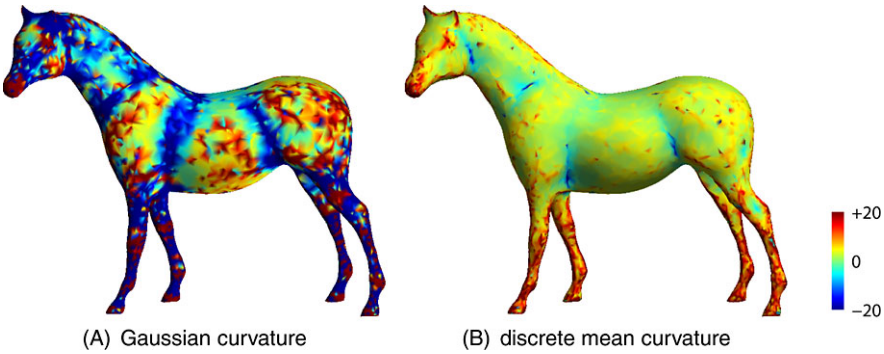


Fig. 8.5 Example of discrete curvature measures applied to horse triangular mesh. (**A**) Gaussian curvature calculated from the Gauss–Bonnet theorem. (**B**) Discrete mean curvature computed from the method of Meyer et al. [280]. Both measures calculate curvature as a node quantity measured for each vertex in the polygonal mesh based on the embedding of the neighboring vertices and the incident faces. Both curvatures are visualized with a common color scale provided on the *lower right*

angle sum, holds for finer and finer mesh spacing and is equivalent to the continuous definition in the limit. In contrast, mean curvature does not possess an analogous integral definition and therefore it is less natural to express mean curvature in the discrete setting. One approach to defining mean curvature, adopted by Meyer et al. [280], is to establish a similar integral relationship in the discrete setting that approaches the continuous definition in the limit of finer mesh sampling in order to provide a robust and compatible curvature measure. The definition begins by noting that if one establishes a specific coordinate system for the surface that provides a isothermal parameterization, the mean curvature H is related to the Laplacian operator ∇^2 applied to this specific choice of coordinate functions [110], and therefore the integral of the mean curvature around a vertex v_i can be expressed as the integral of the Laplacian operator evaluated at the midpoints of all edges incident to v_i. The resulting expression for mean curvature vector is then given by

$$\tilde{H}(v_i) = \frac{1}{2A_{v_i}} \sum_{j \forall e_{ij}} (\cot \alpha_{ij} + \cot \beta_{ij})(\tilde{c}_i - \tilde{c}_j) \tag{8.30}$$

where α_{ij} and β_{ij} represent the opposing angles in the two triangles containing edge e_{ij} at the two vertices that are neither v_i nor v_j. The requirement of the two coordinate vectors $\tilde{c}_i$ and $\tilde{c}_j$ demonstrates why the mean curvature formulation is dependent on the embedding of the graph and is therefore an extrinsic quantity.

Although the mean curvature vector is a vectorial quantity in the ambient embedding space, it is typically expressed as a scalar-valued curvature measure given by the length of the vector, $H = \|\tilde{H}\|$. Intuitively, this definition phrases the mean curvature vector at a vertex as the average edge vector of all edges incident to the vertex weighted by the angle sum around the vertex. Examples of these measures for Gaussian and mean curvature are provided in Fig. 8.5 (contained in the color plate section at the end of the book).

Although these two curvature measures are not phrased in terms of the discrete calculus operators that comprise the central theme of this book, they do provide examples of quantities that are typically considered only in the continuous setting. The definition of Gaussian curvature holds equally well in the discrete setting, whereas the translation of mean curvature to the discrete setting is not as straightforward. From a practical standpoint, these measures are also generally useful for practitioners of discrete methods.

8.5 Applications

Measures of network structure may be useful in several ways. One way to use this information would be to learn more about the network and focus attention on the most relevant nodes. This type of application would employ individual node measures more than the global measures of the network. In this section, we will explore an application in social networks as an illustration of this usage of the network measures.

A different way of using the network measures is to predict other properties of the structure being represented by the network. This type of application effectively summarizes the network structure via a collection of numbers which may then be correlated to other quantities of interest. As a representative of this usage of the network measures, we take an example from chemical graph theory.

8.5.1 Social Networks

Identification of important persons from the network structure is a typical application in a social networks (e.g., [45]). For example, important individuals in a terrorist network could be identified as the most effective leaders to capture. In Chap. 6 we studied Zachary's Karate Club network [415] in which two individuals in the network split the club into two groups. Given just the network structure, we show that the network measures allow us to predict the leaders of the two factions.

Figure 8.6 shows Zachary's Karate Club network with the two actual leaders of the factions identified. Without knowing who the actual leaders of the two factions were, we could examine various node measures. For example, we would expect that important nodes would have a large number of direct social connections to other individuals (nodes) in the network, measured as node degree. Beyond direct connections, we could reasonably expect that the leaders would be well-connected indirectly to all of the individuals (nodes) in the network, measured as a low total distance (closeness). Additionally, we would expect that the leaders would act as a conduit for connecting other individuals in their faction, measured as node betweenness. From Fig. 8.6 we see that the actual leaders of the two factions score much better than the other nodes in terms of the node degree, total distance and betweenness, and are therefore distinct and identifiable from the rest of the nodes.

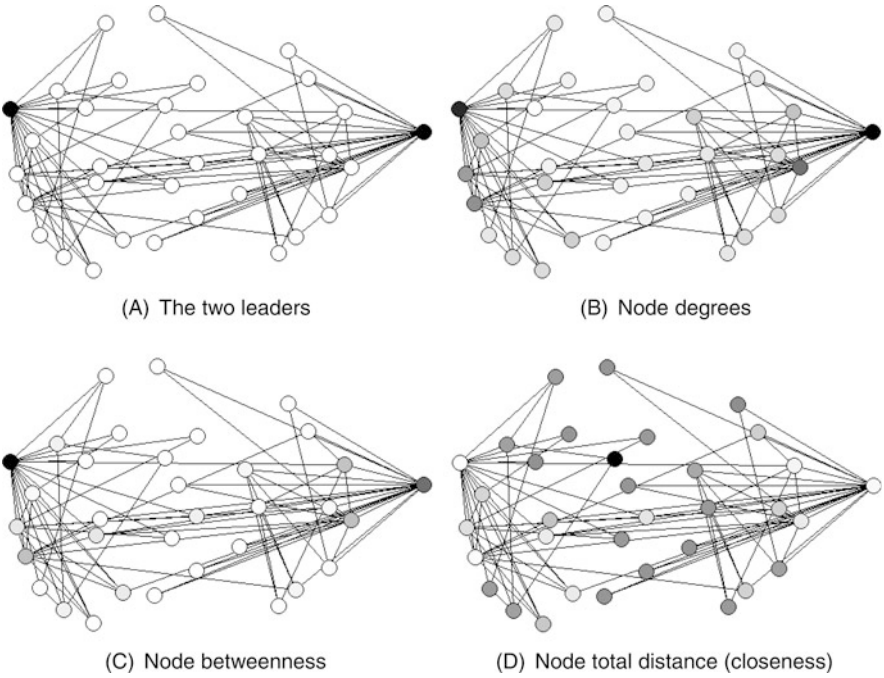


Fig. 8.6 Prediction of the leaders of the two factions in Zachary's Karate Club network [415] using node measures. Nodes are shaded darker to indicate greater values of each measure. Note that the measures in each figure are normalized to use the full white–black range. (**A**) The two actual leaders of the factions marked in *black*. (**B**) Number of neighbors for each node (degree). (**C**) Node betweenness. (**D**) Node total distance (closeness). These measures tell us that the two leaders knew the most other people (highest degrees), were hubs for other people to know each other (highest betweenness) and had the fewest average number of links between them and the rest of the network

Therefore, these three measures of node degree would provide an accurate prediction of the two leaders in Zachary's Karate Club network.

8.5.2 Chemical Graph Theory

Graph theory has a long history in chemistry, dating back to Sylvester's work in 1878 [367]. In fact, the term "graph theory" was coined by Sylvester in the context of the "graphical notation" used to describe the chemical structure of a molecule [37].

In chemistry, a graph represents a molecule by associating each atom with a node and each bond between atoms with an edge. Hydrogen atoms (with just a single bond) are conventionally removed from the graph structure. Network measures are widely used in chemical graph theory to predict **quantitative structure–property relationships** (QSPR) and **quantitative structure–activity relationships**

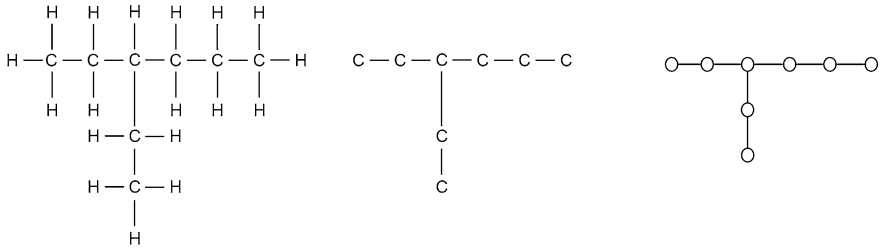


Fig. 8.7 Predicting the boiling point of 3-Ethylhexane from its structure. (**A**) 3-Ethylhexane. (**B**) The hydrogen-depleted molecule commonly used for calculation (i.e., the hydrogen nodes are removed). (**C**) The graph corresponding to the hydrogen-depleted molecule. $W(\mathcal{G}) = 72$, $P = 7$, $\alpha = 1.53$, $\beta = 5.5$, $\gamma = -30.35$. Using Wiener's formula in (8.31) we can predict the boiling point for 3-Ethylhexane as 118.4°C. The actual boiling point for 3-Ethylhexane is 118.6°C

(QSAR). Examples of such molecular properties include the boiling point, melting point, molar volume, refractive index, critical pressure, surface tension, viscosity, rate of electro-reduction and heats of isomerization, vaporization, formation and atomization [109]. Several review articles [187, 282] and books [42, 44, 170] detail the use of these measures in chemical graph theory to predict molecular properties. In addition to molecular properties, network measures have been widely used to predict pharmacological properties for purposes of drug discovery [265, 266, 325].

In this section, we do not intend to provide a comprehensive review of this vast literature, but rather to focus on the first and most important measure to appear in chemical graph theory: the Wiener index (8.6). Note that Wiener's initial definition was not phrased in graph theory language, but later formulated in this setting by Hosoya [207]. In his early papers [404, 405], Wiener used his index to predict the boiling point (b.p.) of alkanes from the formula

$$\text{b.p.} = \alpha W(\mathcal{G}) + \beta P + \gamma, \tag{8.31}$$

where α, β, γ are empirical constants and P, the "polarity number", was defined as the number of node pairs with distance equal to three. By fitting parameters to thirty-seven alkanes, Wiener determined the parameters to be $\alpha = 98/n^2$, $\beta = 5.5$ and $\gamma = -30.35$, where n represents the number of carbon atoms (number of nodes). It is interesting to observe that the only difference between the definitions of the Wiener index (8.6) and the Average Path Length (8.7) was the normalization of the Average Path Length measure by the number of node pairs, $n^2 - n$. However, in Wiener's original work, he used the parameter setting to normalize his measure by n^2, which brings these two measures very close together.

Figure 8.7 shows one of Wiener's original examples in which he predicts the boiling point of 3-Ethylhexane from the molecular structure alone (i.e., by measuring network properties). Specifically, for 3-Ethylhexane we may calculate $W(\mathcal{G}) = 72$, $P = 7$. Using Wiener's formula (8.31) and the parameter values listed above allows us to calculate the boiling point for 3-Ethylhexane as 118.4°C. The actual boiling

point for 3-Ethylhexane is 118.6°C. In Wiener's original paper, he applied his formula to ninety-four compounds to predict the actual boiling points with an average deviation of 0.97°C. It remains a remarkable fact that a measure of the structure of the network representing a molecule can provide such accurate predictions of the molecule's chemical properties.

Many additional models have been developed that predict molecular properties from the Wiener index. Another example is the chromatographic retention time (CRT) of monoalkyl- and o-dialkylbenzenes which are well modeled in [51] by

$$\mathrm{CRT} = \alpha W(\mathcal{G})^{\beta} + \gamma, \tag{8.32}$$

where α, β, γ are empirical constants (which are different from those appearing above in (8.31)). The variety of models in which the Wiener index appears has been explained by arguing that the Wiener index measures the van der Waals surface area of a molecule [181].

By representing a molecule as a list of numerical descriptors, these descriptors may be used in conjunction with machine learning techniques to predict many different QSPR and QSAR properties. In the present day, vast libraries of graph representations of chemical compounds have been compiled which make it possible to search for a compound with a particular set of properties or to numerically screen compounds without having to manufacture and test them [171, 213, 370]. Many structural descriptors have been devised in the chemical graph theory literature that were not reviewed in this chapter. In this section, we have simply intended to provide the reader with a glimpse into this rich literature and to point the interested reader to more comprehensive sources in this area.

8.6 Conclusion

In this chapter we considered methods of measuring different kinds of quantities that describe the structure of the graph. The distance and clustering measures approached the related intuitive concepts of connectedness and separability. In contrast, the topological and geometric measures probe other aspects of the complex.

The measures may be applied in several ways. One method for applying these measures has been to reduce a network to a series of numbers that may be used to predict the behavior of certain processes on the graph or of the object represented by the network. A central example of this usage is the success of the chemical graph theory literature in relating the distance-based measures to the chemical properties of the molecules represented by the graph. A smaller example of this usefulness is the observation [175] that graph diameter is a good predictor of the convergence of conjugate gradient applied to solving a linear system with the Laplacian matrix. As network models are increasingly explored to describe computer networks, neural connections, traffic flow, gene regulation and sociology, we believe that these measurements will provide useful predictions about the behavior of these networks.

Finally we note that the distance and separability descriptors that have been employed in the literature to describe networks are exclusively dependent on node connectivity and separability. We suggested some possibilities for extending these measures to higher-order connectivity, but we believe that measures defined on edge and cycle connectivity and separability present an untapped source of additional descriptors for a network or complex.