# MAT 280: Harmonic Analysis on Graphs & Networks
# Lecture 5: Graph Laplacian Eigenvalues II

*Naoki Saito*

Department of Mathematics
University of California, Davis

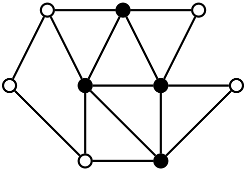
October 10, 2019

# Outline

# Outline

# Isoperimetric Number

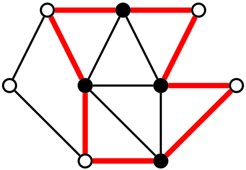
- Let $S \subset V(G)$ be a nonempty subset of vertices of $G$.
- $\partial S := \{e = (u, v) \in E(G) \mid u \in S, v \notin S\}$, which is called the *boundary* of $S$.



$S = \{\bullet\}$, $S^c = \{\circ\}$

# Isoperimetric Number

- Let $S \subset V(G)$ be a nonempty subset of vertices of $G$.
- $\partial S := \{e = (u, v) \in E(G) \mid u \in S, v \notin S\}$, which is called the *boundary* of $S$.



$S = \{\bullet\}$, $S^c = \{\circ\}$

# Isoperimetric Number ...

- The *Cheeger ratio* for $S \subset V$ is defined as

$$h(S) := \frac{|\partial S|}{\min\{|S|, |S^c|\}}.$$

- For the example graph in the previous page: $|S| = 4$; $|S^c| = 5$; $|\partial S| = 8$. Hence, $h(S) = 8/4 = 2$.
- The Cheeger ratio tells us the *quality of the cut* of $V$ into $S \cup S^c$: if $S$ and $S^c$ are well-balanced, i.e., $|S| \approx |S^c|$, and $\exists$ few edges connecting $S$ and $S^c$, then $h(S)$ is small.
- The *isoperimetric number* (or a.k.a. the *Cheeger constant*) $i(G)$ of $G$ is defined as

$$i(G) := \inf_{S \subset V; S \neq \emptyset} \frac{|\partial S|}{\min\{|S|, |S^c|\}}.$$

- This definition is exactly the same as the one given in my previous lecture, but is in a more symmetric form.
- The version for $L_{\text{rw}}$ replaces $|S|$ and $|S^c|$ by $\text{vol}(S)$ and $\text{vol}(S^c)$ where $\text{vol}(S) := \sum_{v \in S} d(v)$. That version has also been studied extensively.

# Isoperimetric Number . . .

- The *Cheeger ratio* for $S \subset V$ is defined as
$$h(S) := \frac{|\partial S|}{\min\{|S|, |S^c|\}}.$$
- For the example graph in the previous page: $|S| = 4$; $|S^c| = 5$; $|\partial S| = 8$. Hence, $h(S) = 8/4 = 2$.
- The Cheeger ratio tells us the *quality of the cut* of $V$ into $S \cup S^c$: if $S$ and $S^c$ are well-balanced, i.e., $|S| \approx |S^c|$, and $\exists$ few edges connecting $S$ and $S^c$, then $h(S)$ is small.
- The *isoperimetric number* (or a.k.a. the *Cheeger constant*) $i(G)$ of $G$ is defined as
$$i(G) := \inf_{S \subset V; S \neq \emptyset} \frac{|\partial S|}{\min\{|S|, |S^c|\}}.$$
- This definition is exactly the same as the one given in my previous lecture, but is in a more symmetric form.
- The version for $L_{\mathrm{rw}}$ replaces $|S|$ and $|S^c|$ by $\mathrm{vol}(S)$ and $\mathrm{vol}(S^c)$ where $\mathrm{vol}(S) := \sum_{v \in S} d(v)$. That version has also been studied extensively.

# Isoperimetric Number ...

- The *Cheeger ratio* for $S \subset V$ is defined as

$$h(S) := \frac{|\partial S|}{\min\{|S|, |S^c|\}}.$$

- For the example graph in the previous page: $|S| = 4$; $|S^c| = 5$; $|\partial S| = 8$. Hence, $h(S) = 8/4 = 2$.
- The Cheeger ratio tells us the *quality of the cut* of $V$ into $S \cup S^c$: if $S$ and $S^c$ are well-balanced, i.e., $|S| \approx |S^c|$, and $\exists$ few edges connecting $S$ and $S^c$, then $h(S)$ is small.
- The *isoperimetric number* (or a.k.a. the *Cheeger constant*) $i(G)$ of $G$ is defined as

$$i(G) := \inf_{S \subset V; S \neq \emptyset} \frac{|\partial S|}{\min\{|S|, |S^c|\}}.$$

- This definition is exactly the same as the one given in my previous lecture, but is in a more symmetric form.
- The version for $L_{\mathrm{rw}}$ replaces $|S|$ and $|S^c|$ by $\mathrm{vol}(S)$ and $\mathrm{vol}(S^c)$ where $\mathrm{vol}(S) := \sum_{v \in S} d(v)$. That version has also been studied extensively.

# Isoperimetric Number . . .

- The *Cheeger ratio* for $S \subset V$ is defined as

$$h(S) := \frac{|\partial S|}{\min\{|S|, |S^c|\}}.$$

- For the example graph in the previous page: $|S| = 4$; $|S^c| = 5$; $|\partial S| = 8$. Hence, $h(S) = 8/4 = 2$.
- The Cheeger ratio tells us the *quality of the cut* of $V$ into $S \cup S^c$: if $S$ and $S^c$ are well-balanced, i.e., $|S| \approx |S^c|$, and $\exists$ few edges connecting $S$ and $S^c$, then $h(S)$ is small.
- The *isoperimetric number* (or a.k.a. the *Cheeger constant*) $i(G)$ of $G$ is defined as

$$i(G) := \inf_{S \subset V; S \neq \emptyset} \frac{|\partial S|}{\min\{|S|, |S^c|\}}.$$

- This definition is exactly the same as the one given in my previous lecture, but is in a more symmetric form.
- The version for $L_{\mathrm{rw}}$ replaces $|S|$ and $|S^c|$ by $\mathrm{vol}(S)$ and $\mathrm{vol}(S^c)$ where $\mathrm{vol}(S) := \sum_{v \in S} d(v)$. That version has also been studied extensively.

# Isoperimetric Number ...

- The *Cheeger ratio* for $S \subset V$ is defined as
$$h(S) := \frac{|\partial S|}{\min\{|S|, |S^c|\}}.$$
- For the example graph in the previous page: $|S| = 4$; $|S^c| = 5$; $|\partial S| = 8$. Hence, $h(S) = 8/4 = 2$.
- The Cheeger ratio tells us the *quality of the cut* of $V$ into $S \cup S^c$: if $S$ and $S^c$ are well-balanced, i.e., $|S| \approx |S^c|$, and $\exists$ few edges connecting $S$ and $S^c$, then $h(S)$ is small.
- The *isoperimetric number* (or a.k.a. the *Cheeger constant*) $i(G)$ of $G$ is defined as
$$i(G) := \inf_{S \subset V; S \neq \emptyset} \frac{|\partial S|}{\min\{|S|, |S^c|\}}.$$
- This definition is exactly the same as the one given in my previous lecture, but is in a more symmetric form.
- The version for $L_{\mathrm{rw}}$ replaces $|S|$ and $|S^c|$ by $\mathrm{vol}(S)$ and $\mathrm{vol}(S^c)$ where $\mathrm{vol}(S) := \sum_{v \in S} d(v)$. That version has also been studied extensively.

# Isoperimetric Number ...

- The *Cheeger ratio* for $S \subset V$ is defined as
$$h(S) := \frac{|\partial S|}{\min\{|S|, |S^c|\}}.$$
- For the example graph in the previous page: $|S| = 4$; $|S^c| = 5$; $|\partial S| = 8$. Hence, $h(S) = 8/4 = 2$.
- The Cheeger ratio tells us the *quality of the cut* of $V$ into $S \cup S^c$: if $S$ and $S^c$ are well-balanced, i.e., $|S| \approx |S^c|$, and ∃ few edges connecting $S$ and $S^c$, then $h(S)$ is small.
- The *isoperimetric number* (or a.k.a. the *Cheeger constant*) $i(G)$ of $G$ is defined as
$$i(G) := \inf_{S \subset V; S \neq \emptyset} \frac{|\partial S|}{\min\{|S|, |S^c|\}}.$$
- This definition is exactly the same as the one given in my previous lecture, but is in a more symmetric form.
- The version for $L_{\mathrm{rw}}$ replaces $|S|$ and $|S^c|$ by $\mathrm{vol}(S)$ and $\mathrm{vol}(S^c)$ where $\mathrm{vol}(S) := \sum_{v \in S} d(v)$. That version has also been studied extensively.

# Why is $i(G)$ so important or interesting?

- To determine it, we need to find a small *edge-cut* separating as large a subset $S$ with $|S| \leq n/2$ as possible from the remaining larger part $S^c$. $\Longrightarrow$ It serves as a *measure of connectivity* of $G$. May indicate how easy is it to "destroy" a given network $G$ by cutting only a few edges.
- The problem of partitioning $V(G)$ into two equally sized subsets (to within one element) in such a way that the number of the edges in the cut is minimal, is known as the *bisection width* problem. There are many practical applications, e.g., VLSI design, etc.
- $i(G)$: large $\Longrightarrow$ $G$ has a large *growth rate*. More precisely, let $B_k(v)$ be the set of vertices of $G$ at distance at most $k$ from $v$, like a ball with center $v$ and radius $k$. Then, $|B_{k+1}(v)|/|B_k(v)| \geq i(G)/d_{\max}(G)$.
- It is a discrete analogue of the *Cheeger constant* in Riemannian geometry, i.e., the minimal area of a hypersurface that divides a given compact Riemannian manifold into two disjoint pieces of equal volume.

# Why is $i(G)$ so important or interesting?

- To determine it, we need to find a small *edge-cut* separating as large a subset $S$ with $|S| \le n/2$ as possible from the remaining larger part $S^c$. $\Longrightarrow$ It serves as a *measure of connectivity* of $G$. May indicate how easy is it to "destroy" a given network $G$ by cutting only a few edges.
- The problem of partitioning $V(G)$ into two equally sized subsets (to within one element) in such a way that the number of the edges in the cut is minimal, is known as the *bisection width* problem. There are many practical applications, e.g., VLSI design, etc.
- $i(G)$: large $\Longrightarrow$ $G$ has a large *growth rate*. More precisely, let $B_k(v)$ be the set of vertices of $G$ at distance at most $k$ from $v$, like a ball with center $v$ and radius $k$. Then, $|B_{k+1}(v)|/|B_k(v)| \ge i(G)/d_{\max}(G)$.
- It is a discrete analogue of the *Cheeger constant* in Riemannian geometry, i.e., the minimal area of a hypersurface that divides a given compact Riemannian manifold into two disjoint pieces of equal volume.

# Why is $i(G)$ so important or interesting?

- To determine it, we need to find a small *edge-cut* separating as large a subset $S$ with $|S| \le n/2$ as possible from the remaining larger part $S^c$. $\Longrightarrow$ It serves as a *measure of connectivity* of $G$. May indicate how easy is it to "destroy" a given network $G$ by cutting only a few edges.
- The problem of partitioning $V(G)$ into two equally sized subsets (to within one element) in such a way that the number of the edges in the cut is minimal, is known as the *bisection width* problem. There are many practical applications, e.g., VLSI design, etc.
- $i(G)$: large $\Longrightarrow$ $G$ has a large *growth rate*. More precisely, let $B_k(v)$ be the set of vertices of $G$ at distance at most $k$ from $v$, like a ball with center $v$ and radius $k$. Then, $|B_{k+1}(v)|/|B_k(v)| \ge i(G)/d_{\max}(G)$.
- It is a discrete analogue of the *Cheeger constant* in Riemannian geometry, i.e., the minimal area of a hypersurface that divides a given compact Riemannian manifold into two disjoint pieces of equal volume.

# Why is $i(G)$ so important or interesting?

- To determine it, we need to find a small *edge-cut* separating as large a subset $S$ with $|S| \le n/2$ as possible from the remaining larger part $S^c$. $\Longrightarrow$ It serves as a *measure of connectivity* of $G$. May indicate how easy is it to "destroy" a given network $G$ by cutting only a few edges.
- The problem of partitioning $V(G)$ into two equally sized subsets (to within one element) in such a way that the number of the edges in the cut is minimal, is known as the *bisection width* problem. There are many practical applications, e.g., VLSI design, etc.
- $i(G)$: large $\Longrightarrow$ $G$ has a large *growth rate*. More precisely, let $B_k(v)$ be the set of vertices of $G$ at distance at most $k$ from $v$, like a ball with center $v$ and radius $k$. Then, $|B_{k+1}(v)|/|B_k(v)| \ge i(G)/d_{\max}(G)$.
- It is a discrete analogue of the *Cheeger constant* in Riemannian geometry, i.e., the minimal area of a hypersurface that divides a given compact Riemannian manifold into two disjoint pieces of equal volume.

# Isoperimetric Numbers of Specific Graphs

| Graph $G$ | Isoperimetric Number $i(G)$ |
|---|---|
| Complete graph $K_n$ | $i(K_n) = \lceil n/2 \rceil$ |
| Path $P_n$ | $i(P_n) = 1/\lfloor n/2 \rfloor$ |
| Cycle $C_n$ | $i(C_n) = 2/\lfloor n/2 \rfloor$ |
| Bipartite complete graph $K_{p,q}$ | $i\left(K_{p,q}\right) = \lceil pq/2 \rceil / \lfloor (p+q)/2 \rfloor$ |
| Cube $m$-dimension $Cb_m$ | $i(Cb_m) = 1$ |
| Petersen Graph $P$ | $i(P) = 1$ |

# The Isoperimetric Number and Algebraic Connectivity

- Both are viewed as measures of connectivity of a given graph.
- While $i(G)$ is more explicitly related to the connectivity of a graph than $a(G)$, it is more difficult (i.e., combinatorial) to compute $i(G)$ than $a(G)$.
- Hence, the bounds of $i(G)$ in terms of $a(G)$ and the other quantities have been extensively studied.
- Mohar (1987, 1989): $a(G)/2 \le i(G) \lneq \sqrt{a(G)\,(2d_{\max}(G) - a(G))}$ for $n \ge 4$, where $d_{\max}(G) := \max_j d_j$.

# The Isoperimetric Number and Algebraic Connectivity

- Both are viewed as measures of connectivity of a given graph.
- While $i(G)$ is more explicitly related to the connectivity of a graph than $a(G)$, it is more difficult (i.e., combinatorial) to compute $i(G)$ than $a(G)$.
- Hence, the bounds of $i(G)$ in terms of $a(G)$ and the other quantities have been extensively studied.
- Mohar (1987, 1989): $a(G)/2 \le i(G) \lneqq \sqrt{a(G)\,(2d_{\max}(G) - a(G))}$ for $n \ge 4$, where $d_{\max}(G) := \max_j d_j$.

# The Isoperimetric Number and Algebraic Connectivity

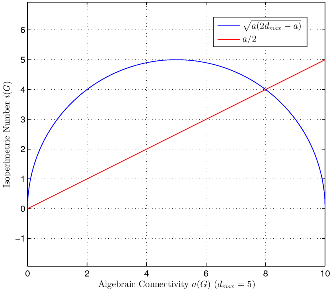
- Both are viewed as measures of connectivity of a given graph.
- While $i(G)$ is more explicitly related to the connectivity of a graph than $a(G)$, it is more difficult (i.e., combinatorial) to compute $i(G)$ than $a(G)$.
- Hence, the bounds of $i(G)$ in terms of $a(G)$ and the other quantities have been extensively studied.
- Mohar (1987, 1989): $a(G)/2 \le i(G) \lneqq \sqrt{a(G)\,(2d_{\max}(G) - a(G))}$ for $n \ge 4$, where $d_{\max}(G) := \max_j d_j$.

# The Isoperimetric Number and Algebraic Connectivity

- Both are viewed as measures of connectivity of a given graph.
- While $i(G)$ is more explicitly related to the connectivity of a graph than $a(G)$, it is more difficult (i.e., combinatorial) to compute $i(G)$ than $a(G)$.
- Hence, the bounds of $i(G)$ in terms of $a(G)$ and the other quantities have been extensively studied.
- Mohar (1987, 1989): $a(G)/2 \le i(G) \lneqq \sqrt{a(G)\,(2d_{\max}(G) - a(G))}$ for $n \ge 4$, where $d_{\max}(G) := \max_j d_j$.

# The Isoperimetric Number and Algebraic Connectivity

- Both are viewed as measures of connectivity of a given graph.
- While $i(G)$ is more explicitly related to the connectivity of a graph than $a(G)$, it is more difficult (i.e., combinatorial) to compute $i(G)$ than $a(G)$.
- Hence, the bounds of $i(G)$ in terms of $a(G)$ and the other quantities have been extensively studied.
- Mohar (1987, 1989): $a(G)/2 \le i(G) \lneqq \sqrt{a(G)\,(2d_{\max}(G) - a(G))}$ for $n \ge 4$, where $d_{\max}(G) := \max_i d_i$.



# Outline

# Isospectrality

- The spectrum (i.e., the set of eigenvalues) $\Lambda(G)$ of $L(G)$ cannot uniquely determine the graph $G$.

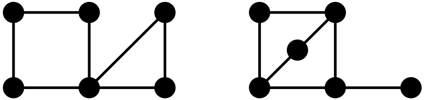
  $\sim$ Kac (1966): "Can one hear the shape of a drum?" $\Longrightarrow$ Gordon, Webb, & Wolpert (1992): "One cannot hear the shape of a drum."
- An example of "isospectral" graphs (Tan, 1998; Fujii & Katsuda, 1999):

# Isospectrality

- The spectrum (i.e., the set of eigenvalues) $\Lambda(G)$ of $L(G)$ cannot uniquely determine the graph $G$.
  $\sim$ Kac (1966): "Can one hear the shape of a drum?" $\Longrightarrow$ Gordon, Webb, & Wolpert (1992): "One cannot hear the shape of a drum."
- An example of "isospectral" graphs (Tan, 1998; Fujii & Katsuda, 1999):

# Isospectrality

- The spectrum (i.e., the set of eigenvalues) $\Lambda(G)$ of $L(G)$ cannot uniquely determine the graph $G$.
  ∼ Kac (1966): "Can one hear the shape of a drum?" $\Longrightarrow$ Gordon, Webb, & Wolpert (1992): "One cannot hear the shape of a drum."
- An example of "isospectral" graphs (Tan, 1998; Fujii & Katsuda, 1999):



$$L(G_1) = \begin{bmatrix} 2 & -1 & 0 & 0 & 0 & -1 \\ -1 & 2 & -1 & 0 & 0 & 0 \\ 0 & -1 & 4 & -1 & -1 & -1 \\ 0 & 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & -1 & -1 & 2 & 0 \\ -1 & 0 & -1 & 0 & 0 & 2 \end{bmatrix} \neq L(G_2) = \begin{bmatrix} 2 & -1 & 0 & 0 & -1 & 0 \\ -1 & 3 & -1 & 0 & 0 & -1 \\ 0 & -1 & 3 & -1 & -1 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 \\ -1 & 0 & -1 & 0 & 3 & -1 \\ 0 & -1 & 0 & 0 & -1 & 2 \end{bmatrix}$$

But, $\Lambda(G_1) = \Lambda(G_2) = \{0, 0.7639, 2, 3, 3, 5.2361\}$.

# Spectral Characterization of Certain Classes of Graphs

- In fact, *there are 58 pairs, 6 triples of isospectral graphs within all possible simple/undirected/unweighted graphs with* $n < 8$ (Tan, 1998).
- However, certain classes of graphs can be completely determined by their Laplacian spectra: starlike trees (Omidi & Tajbakhsh, 2007), centipedes (Boulet, 2008), ...
- ∃ some attempts to reconstruct graphs from their Laplacian spectra via combinatorial optimization (e.g., Comellas & Diaz-Lopez, 2008)
- Nothing prevents us from using the Laplacian spectra for characterizing dendrite patterns!

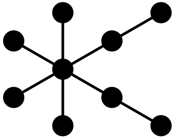
# Spectral Characterization of Certain Classes of Graphs

- In fact, *there are 58 pairs, 6 triples of isospectral graphs within all possible simple/undirected/unweighted graphs with $n < 8$* (Tan, 1998).
- However, certain classes of graphs can be completely determined by their Laplacian spectra: starlike trees (Omidi & Tajbakhsh, 2007), centipedes (Boulet, 2008), ...

- ∃ some attempts to reconstruct graphs from their Laplacian spectra via combinatorial optimization (e.g., Comellas & Diaz-Lopez, 2008)
- Nothing prevents us from using the Laplacian spectra for characterizing dendrite patterns!

# Spectral Characterization of Certain Classes of Graphs

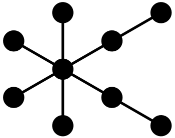
- In fact, *there are 58 pairs, 6 triples of isospectral graphs within all possible simple/undirected/unweighted graphs with* $n < 8$ (Tan, 1998).
- However, certain classes of graphs can be completely determined by their Laplacian spectra: starlike trees (Omidi & Tajbakhsh, 2007), centipedes (Boulet, 2008), ...

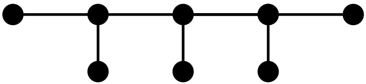


- ∃ some attempts to reconstruct graphs from their Laplacian spectra via combinatorial optimization (e.g., Comellas & Diaz-Lopez, 2008)
- Nothing prevents us from using the Laplacian spectra for characterizing dendrite patterns!

# Spectral Characterization of Certain Classes of Graphs

- In fact, *there are 58 pairs, 6 triples of isospectral graphs within all possible simple/undirected/unweighted graphs with $n < 8$* (Tan, 1998).
- However, certain classes of graphs can be completely determined by their Laplacian spectra: starlike trees (Omidi & Tajbakhsh, 2007), centipedes (Boulet, 2008), ...

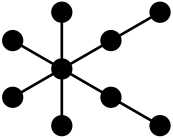


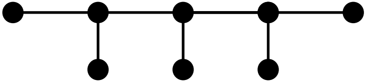


- ∃ some attempts to reconstruct graphs from their Laplacian spectra via combinatorial optimization (e.g., Comellas & Diaz-Lopez, 2008)
- Nothing prevents us from using the Laplacian spectra for characterizing dendrite patterns!

# Spectral Characterization of Certain Classes of Graphs

- In fact, *there are 58 pairs, 6 triples of isospectral graphs within all possible simple/undirected/unweighted graphs with $n < 8$* (Tan, 1998).
- However, certain classes of graphs can be completely determined by their Laplacian spectra: starlike trees (Omidi & Tajbakhsh, 2007), centipedes (Boulet, 2008), ...

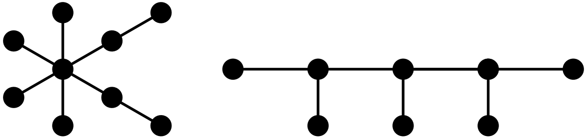




- ∃ some attempts to reconstruct graphs from their Laplacian spectra via combinatorial optimization (e.g., Comellas & Diaz-Lopez, 2008)
- Nothing prevents us from using the Laplacian spectra for characterizing dendrite patterns!

# Spectral Characterization of Certain Classes of Graphs

- In fact, *there are 58 pairs, 6 triples of isospectral graphs within all possible simple/undirected/unweighted graphs with $n < 8$* (Tan, 1998).
- However, certain classes of graphs can be completely determined by their Laplacian spectra: starlike trees (Omidi & Tajbakhsh, 2007), centipedes (Boulet, 2008), . . .

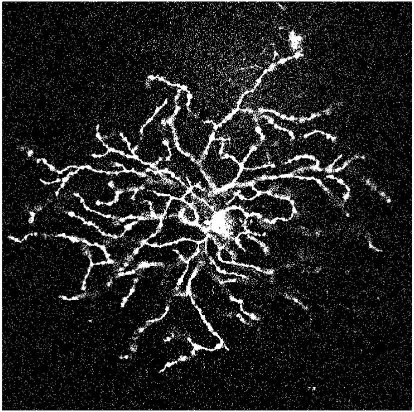


- ∃ some attempts to reconstruct graphs from their Laplacian spectra via combinatorial optimization (e.g., Comellas & Diaz-Lopez, 2008)
- Nothing prevents us from using the Laplacian spectra for characterizing dendrite patterns!

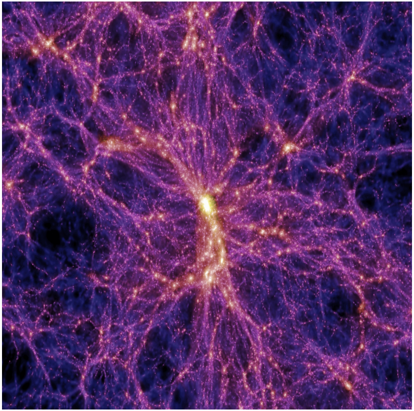
# Outline

# Outline

# Morphology of Network-like Structures

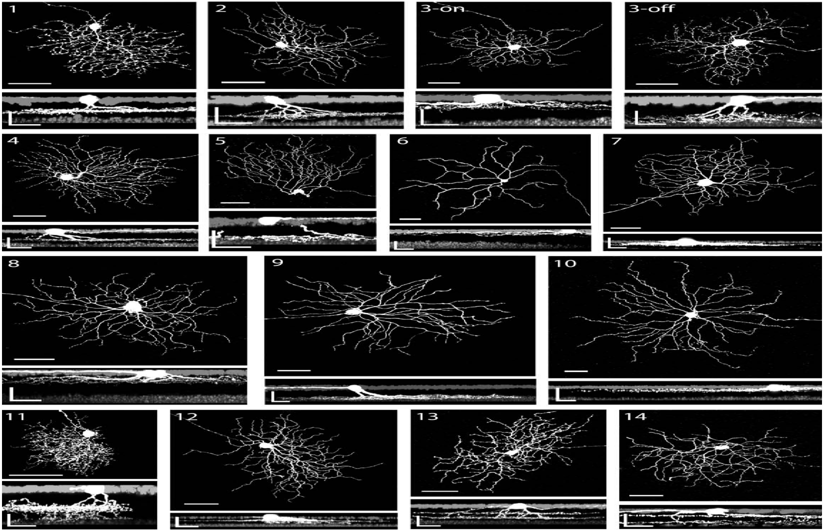


(a) Neuron

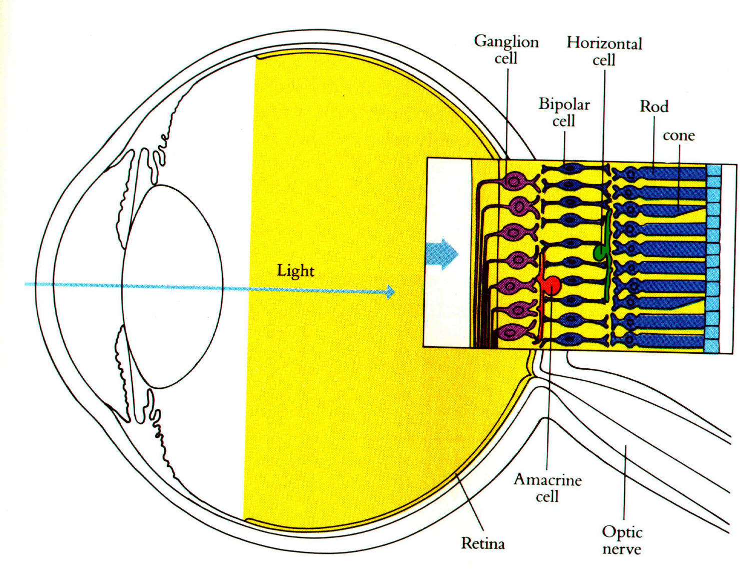


(b) Universe

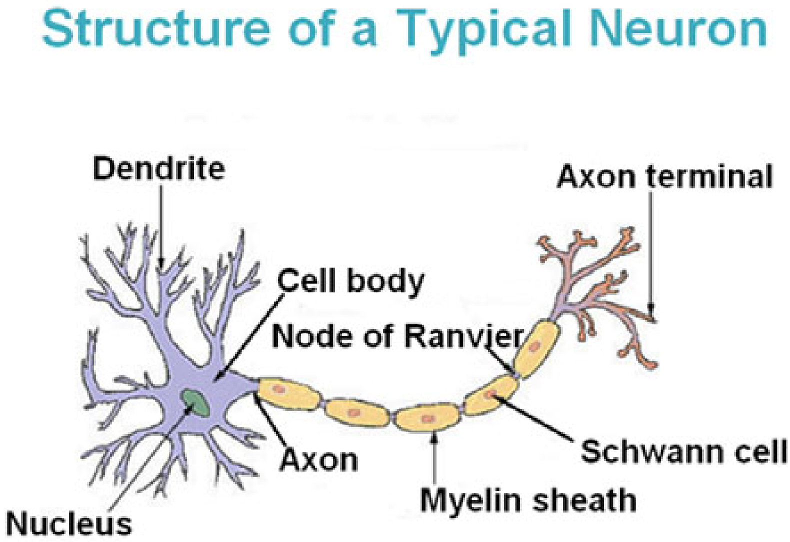
# Morphology of Retinal Ganglion Cells



# Retinal Ganglion Cells (D. Hubel: *Eye, Brain, & Vision*, '95)



# A Typical Neuron (from Wikipedia)



# Clustering Mouse's Retinal Ganglion Cells

- Neuroscientists' Objective: To understand how *structural* / *morphological* properties of dendritic trees of mouse retinal ganglion cells (RGCs) relate to the *cell types* and their *functionality*; how such properties *change* / *evolve* from newborn to adult
- Why mouse? $\Longrightarrow$ Great possibilities for genetic manipulation
- Data: 3D images of dendrites of RGCs via a confocal microscope
- State of the art: A manually intensive procedure using specialized software[1]:
  - Trace and segment dendrite patterns from each 3D cube;
  - Extract geometric/morphological parameters (totally 14 parameters);
  - Apply a conventional bottom-up "hierarchical clustering" algorithm
- The extracted morphological parameters include: somal size; dendritic field size; total dendrite length; branch order; mean internal branch length; branch angle; mean terminal branch length, ...
- It takes *half a day per cell with a lot of human interactions!*

[1] Neurolucida®, MBF Bioscience

# Clustering Mouse's Retinal Ganglion Cells

- Neuroscientists' Objective: To understand how *structural* / *morphological* properties of dendritic trees of mouse retinal ganglion cells (RGCs) relate to the *cell types* and their *functionality*; how such properties *change* / *evolve* from newborn to adult
- Why mouse? $\Longrightarrow$ Great possibilities for genetic manipulation
- Data: 3D images of dendrites of RGCs via a confocal microscope
- State of the art: A manually intensive procedure using specialized software[1]:
  - Trace and segment dendrite patterns from each 3D cube;
  - Extract geometric/morphological parameters (totally 14 parameters);
  - Apply a conventional bottom-up "hierarchical clustering" algorithm
- The extracted morphological parameters include: somal size; dendritic field size; total dendrite length; branch order; mean internal branch length; branch angle; mean terminal branch length, . . .
- It takes *half a day per cell with a lot of human interactions!*

[1] Neurolucida®, MBF Bioscience

# Clustering Mouse's Retinal Ganglion Cells

- Neuroscientists' Objective: To understand how *structural* / *morphological* properties of dendritic trees of mouse retinal ganglion cells (RGCs) relate to the *cell types* and their *functionality*; how such properties *change* / *evolve* from newborn to adult
- Why mouse? $\Longrightarrow$ Great possibilities for genetic manipulation
- Data: 3D images of dendrites of RGCs via a confocal microscope
- State of the art: A manually intensive procedure using specialized software[1]:
  - Trace and segment dendrite patterns from each 3D cube;
  - Extract geometric/morphological parameters (totally 14 parameters);
  - Apply a conventional bottom-up "hierarchical clustering" algorithm
- The extracted morphological parameters include: somal size; dendritic field size; total dendrite length; branch order; mean internal branch length; branch angle; mean terminal branch length, ...
- It takes *half a day per cell with a lot of human interactions!*

[1] Neurolucida®, MBF Bioscience

# Clustering Mouse's Retinal Ganglion Cells

- Neuroscientists' Objective: To understand how *structural* / *morphological* properties of dendritic trees of mouse retinal ganglion cells (RGCs) relate to the *cell types* and their *functionality*; how such properties *change* / *evolve* from newborn to adult
- Why mouse? $\Longrightarrow$ Great possibilities for genetic manipulation
- Data: 3D images of dendrites of RGCs via a confocal microscope
- State of the art: A manually intensive procedure using specialized software[1]:
  - Trace and segment dendrite patterns from each 3D cube;
  - Extract geometric/morphological parameters (totally 14 parameters);
  - Apply a conventional bottom-up "hierarchical clustering" algorithm
- The extracted morphological parameters include: somal size; dendritic field size; total dendrite length; branch order; mean internal branch length; branch angle; mean terminal branch length, . . .
- It takes *half a day per cell with a lot of human interactions!*

[1] Neurolucida®, MBF Bioscience

# Clustering Mouse's Retinal Ganglion Cells

- Neuroscientists' Objective: To understand how *structural* / *morphological* properties of dendritic trees of mouse retinal ganglion cells (RGCs) relate to the *cell types* and their *functionality*; how such properties *change* / *evolve* from newborn to adult
- Why mouse? $\Longrightarrow$ Great possibilities for genetic manipulation
- Data: 3D images of dendrites of RGCs via a confocal microscope
- State of the art: A manually intensive procedure using specialized software[1]:
  - Trace and segment dendrite patterns from each 3D cube;
  - Extract geometric/morphological parameters (totally 14 parameters);
  - Apply a conventional bottom-up "hierarchical clustering" algorithm
- The extracted morphological parameters include: somal size; dendritic field size; total dendrite length; branch order; mean internal branch length; branch angle; mean terminal branch length, ...
- It takes *half a day per cell with a lot of human interactions!*

[1] Neurolucida®, MBF Bioscience

# Clustering Mouse's Retinal Ganglion Cells

- Neuroscientists' Objective: To understand how *structural* / *morphological* properties of dendritic trees of mouse retinal ganglion cells (RGCs) relate to the *cell types* and their *functionality*; how such properties *change* / *evolve* from newborn to adult
- Why mouse? $\Longrightarrow$ Great possibilities for genetic manipulation
- Data: 3D images of dendrites of RGCs via a confocal microscope
- State of the art: A manually intensive procedure using specialized software[1]:
  - Trace and segment dendrite patterns from each 3D cube;
  - Extract geometric/morphological parameters (totally 14 parameters);
  - Apply a conventional bottom-up "hierarchical clustering" algorithm
- The extracted morphological parameters include: somal size; dendritic field size; total dendrite length; branch order; mean internal branch length; branch angle; mean terminal branch length, . . .
- It takes *half a day per cell with a lot of human interactions!*

[1] Neurolucida®, MBF Bioscience

# Clustering Mouse's Retinal Ganglion Cells

- Neuroscientists' Objective: To understand how *structural* / *morphological* properties of dendritic trees of mouse retinal ganglion cells (RGCs) relate to the *cell types* and their *functionality*; how such properties *change* / *evolve* from newborn to adult
- Why mouse? $\Longrightarrow$ Great possibilities for genetic manipulation
- Data: 3D images of dendrites of RGCs via a confocal microscope
- State of the art: A manually intensive procedure using specialized software[1]:
  - Trace and segment dendrite patterns from each 3D cube;
  - Extract geometric/morphological parameters (totally 14 parameters);
  - Apply a conventional bottom-up "hierarchical clustering" algorithm
- The extracted morphological parameters include: somal size; dendritic field size; total dendrite length; branch order; mean internal branch length; branch angle; mean terminal branch length, . . .
- It takes *half a day per cell with a lot of human interactions!*

[1] Neurolucida®, MBF Bioscience

# Clustering Mouse's Retinal Ganglion Cells

- Neuroscientists' Objective: To understand how *structural* / *morphological* properties of dendritic trees of mouse retinal ganglion cells (RGCs) relate to the *cell types* and their *functionality*; how such properties *change* / *evolve* from newborn to adult
- Why mouse? $\Longrightarrow$ Great possibilities for genetic manipulation
- Data: 3D images of dendrites of RGCs via a confocal microscope
- State of the art: A manually intensive procedure using specialized software[1]:
  - Trace and segment dendrite patterns from each 3D cube;
  - Extract geometric/morphological parameters (totally 14 parameters);
  - Apply a conventional bottom-up "hierarchical clustering" algorithm
- The extracted morphological parameters include: somal size; dendritic field size; total dendrite length; branch order; mean internal branch length; branch angle; mean terminal branch length, . . .
- It takes *half a day per cell with a lot of human interactions!*

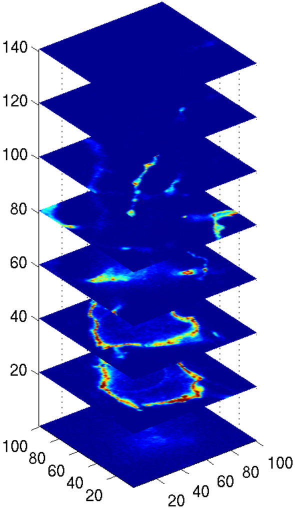
[1] Neurolucida®, MBF Bioscience

# Clustering Mouse's Retinal Ganglion Cells

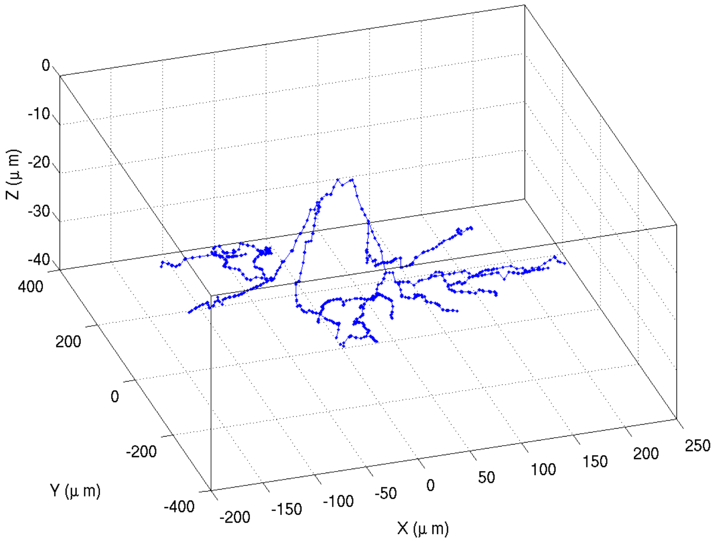
- Neuroscientists' Objective: To understand how *structural* / *morphological* properties of dendritic trees of mouse retinal ganglion cells (RGCs) relate to the *cell types* and their *functionality*; how such properties *change* / *evolve* from newborn to adult
- Why mouse? $\Longrightarrow$ Great possibilities for genetic manipulation
- Data: 3D images of dendrites of RGCs via a confocal microscope
- State of the art: A manually intensive procedure using specialized software[1]:
  - Trace and segment dendrite patterns from each 3D cube;
  - Extract geometric/morphological parameters (totally 14 parameters);
  - Apply a conventional bottom-up "hierarchical clustering" algorithm
- The extracted morphological parameters include: somal size; dendritic field size; total dendrite length; branch order; mean internal branch length; branch angle; mean terminal branch length, ...
- It takes *half a day per cell with a lot of human interactions!*

[1] Neurolucida®, MBF Bioscience

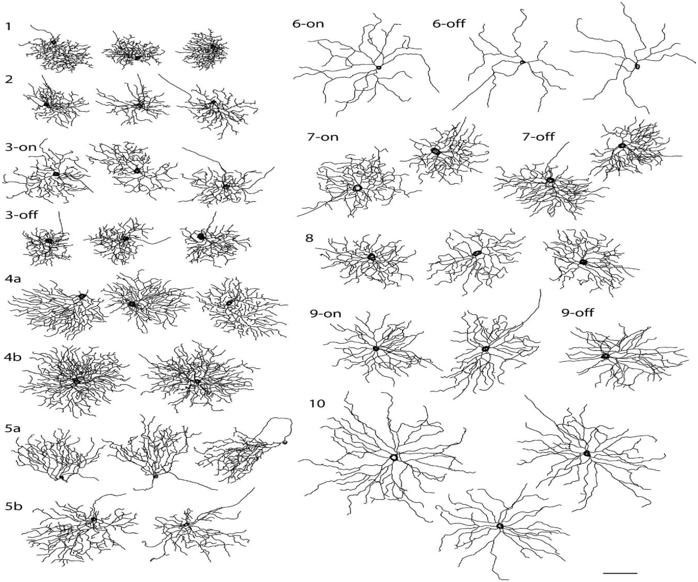
# 3D Data



# Mouse's RGC as a Graph



# Clustering using Features Derived by Neurolucida®



# Our Goal

- We want to develop algorithms for *automatic* morphological feature extraction and clustering from such dendritic trees.
- To do so, we need to convert each dendritic tree to a feature vector in $\mathbb{R}^k$ with relatively small $k \in \mathbb{N}$. This is called a *graph embedding* into a vector space.
- In this lecture, we mainly consider the features using *eigenvalues of graph Laplacians* of such trees.

# Our Goal

- We want to develop algorithms for *automatic* morphological feature extraction and clustering from such dendritic trees.
- To do so, we need to convert each dendritic tree to a feature vector in $\mathbb{R}^k$ with relatively small $k \in \mathbb{N}$. This is called a *graph embedding* into a vector space.
- In this lecture, we mainly consider the features using *eigenvalues of graph Laplacians* of such trees.

# Our Goal

- We want to develop algorithms for *automatic* morphological feature extraction and clustering from such dendritic trees.
- To do so, we need to convert each dendritic tree to a feature vector in $\mathbb{R}^k$ with relatively small $k \in \mathbb{N}$. This is called a *graph embedding* into a vector space.
- In this lecture, we mainly consider the features using *eigenvalues of graph Laplacians* of such trees.

# Our Dataset

consists of 130 RGCs each of which in turn consists of

- A sequence of 3D sample points along dendrite arbors obtained by Neurolucida® (requires intensive human interaction)
- Connectivity and branching information by the same software
- Each soma (cell body) is represented as a sequence of points traced along its boundary (circular/ring shape) $\Longrightarrow$ By replacing such a soma ring by a single vertex representing a center of the soma, each dendritic tree of an RGC is literally represented by a *tree*!
- At this point, we only consider *unweighted* trees.
- $n = |V(G)|$ ranges between 565 and 24474 depending on the RGCs.
- The range of maximum degrees:

$$\max_{130 \text{ cells}} \max_k d(v_k) = 8, \quad \min_{130 \text{ cells}} \max_k d(v_k) = 3.$$

# Our Dataset

consists of 130 RGCs each of which in turn consists of

- A sequence of 3D sample points along dendrite arbors obtained by Neurolucida® (requires intensive human interaction)
- Connectivity and branching information by the same software
- Each soma (cell body) is represented as a sequence of points traced along its boundary (circular/ring shape) $\Longrightarrow$ By replacing such a soma ring by a single vertex representing a center of the soma, each dendritic tree of an RGC is literally represented by a *tree*!
- At this point, we only consider *unweighted* trees.
- $n = |V(G)|$ ranges between 565 and 24474 depending on the RGCs.
- The range of maximum degrees:

$$\max_{130 \text{ cells}} \max_k d(v_k) = 8, \qquad \min_{130 \text{ cells}} \max_k d(v_k) = 3.$$

# Our Dataset

consists of 130 RGCs each of which in turn consists of

- A sequence of 3D sample points along dendrite arbors obtained by Neurolucida® (requires intensive human interaction)
- Connectivity and branching information by the same software
- Each soma (cell body) is represented as a sequence of points traced along its boundary (circular/ring shape) $\Longrightarrow$ By replacing such a soma ring by a single vertex representing a center of the soma, each dendritic tree of an RGC is literally represented by a *tree*!
- At this point, we only consider *unweighted* trees.
- $n = |V(G)|$ ranges between 565 and 24474 depending on the RGCs.
- The range of maximum degrees:

$$\max_{130 \text{ cells}} \max_k d(v_k) = 8, \qquad \min_{130 \text{ cells}} \max_k d(v_k) = 3.$$

# Our Dataset

consists of 130 RGCs each of which in turn consists of

- A sequence of 3D sample points along dendrite arbors obtained by Neurolucida® (requires intensive human interaction)
- Connectivity and branching information by the same software
- Each soma (cell body) is represented as a sequence of points traced along its boundary (circular/ring shape) $\Longrightarrow$ By replacing such a soma ring by a single vertex representing a center of the soma, each dendritic tree of an RGC is literally represented by a *tree*!
- At this point, we only consider *unweighted* trees.
- $n = |V(G)|$ ranges between 565 and 24474 depending on the RGCs.
- The range of maximum degrees:

$$\max_{130 \text{ cells}} \max_k d(v_k) = 8, \qquad \min_{130 \text{ cells}} \max_k d(v_k) = 3.$$

# Our Dataset

consists of 130 RGCs each of which in turn consists of

- A sequence of 3D sample points along dendrite arbors obtained by Neurolucida® (requires intensive human interaction)
- Connectivity and branching information by the same software
- Each soma (cell body) is represented as a sequence of points traced along its boundary (circular/ring shape) $\Longrightarrow$ By replacing such a soma ring by a single vertex representing a center of the soma, each dendritic tree of an RGC is literally represented by a *tree*!
- At this point, we only consider *unweighted* trees.
- $n = |V(G)|$ ranges between 565 and 24474 depending on the RGCs.
- The range of maximum degrees:

$$\max_{130 \text{ cells}} \max_k d(v_k) = 8, \qquad \min_{130 \text{ cells}} \max_k d(v_k) = 3.$$

# Our Dataset

consists of 130 RGCs each of which in turn consists of

- A sequence of 3D sample points along dendrite arbors obtained by Neurolucida® (requires intensive human interaction)
- Connectivity and branching information by the same software
- Each soma (cell body) is represented as a sequence of points traced along its boundary (circular/ring shape) $\implies$ By replacing such a soma ring by a single vertex representing a center of the soma, each dendritic tree of an RGC is literally represented by a *tree*!
- At this point, we only consider *unweighted* trees.
- $n = |V(G)|$ ranges between 565 and 24474 depending on the RGCs.
- The range of maximum degrees:

$$\max_{\text{130 cells}} \max_k d(v_k) = 8, \quad \min_{\text{130 cells}} \max_k d(v_k) = 3.$$

# Outline

# Features Used in Our Experiments

Feature 1: $(p(G) - m_G(1))/|V(G)|$ as a lower bound of the number of pendant neighbors $q(G)$ normalized by $n = |V(G)|$ ;

Feature 2: The normalized Wiener index $W(G)/|V(G)|$ ;

Feature 3: $m_G(4,\infty)/|V(G)|$, i.e., the number of eigenvalues of $L(G)$ larger than 4 (normalized) ;

Feature 4: $\sqrt{a(G)\left(2\max_{v\in V(G)} d_v - a(G)\right)}$, i.e., the upper bound of the isoperimetric number $i(G)$.

- We normalized Features 1, 2, 3, by $n = |V(G)|$ because we wanted to make features less dependent on the number of samples or how the dendrite arbors are sampled. Of course, the number of vertices itself could be a feature although it may not be a decisive one.
- Feature 4 was not explicitly normalized because the isoperimetric number $i(G)$ itself is a normalized quantity in terms of number of vertices.

# Features Used in Our Experiments

Feature 1: $(p(G) - m_G(1))/|V(G)|$ as a lower bound of the number of pendant neighbors $q(G)$ normalized by $n = |V(G)|$ ;

Feature 2: The normalized Wiener index $W(G)/|V(G)|$ ;

Feature 3: $m_G(4,\infty)/|V(G)|$, i.e., the number of eigenvalues of $L(G)$ larger than 4 (normalized) ;

Feature 4: $\sqrt{a(G)\left(2\max_{v \in V(G)} d_v - a(G)\right)}$, i.e., the upper bound of the isoperimetric number $i(G)$.

- We normalized Features 1, 2, 3, by $n = |V(G)|$ because we wanted to make features less dependent on the number of samples or how the dendrite arbors are sampled. Of course, the number of vertices itself could be a feature although it may not be a decisive one.
- Feature 4 was not explicitly normalized because the isoperimetric number $i(G)$ itself is a normalized quantity in terms of number of vertices.

# Features Used in Our Experiments

Feature 1: $(p(G) - m_G(1))/|V(G)|$ as a lower bound of the number of pendant neighbors $q(G)$ normalized by $n = |V(G)|$ ;

Feature 2: The normalized Wiener index $W(G)/|V(G)|$ ;

Feature 3: $m_G(4,\infty)/|V(G)|$, i.e., the number of eigenvalues of $L(G)$ larger than $4$ (normalized) ;

Feature 4: $\sqrt{a(G)\left(2\max_{v\in V(G)} d_v - a(G)\right)}$, i.e., the upper bound of the isoperimetric number $i(G)$.

- We normalized Features 1, 2, 3, by $n = |V(G)|$ because we wanted to make features less dependent on the number of samples or how the dendrite arbors are sampled. Of course, the number of vertices itself could be a feature although it may not be a decisive one.
- Feature 4 was not explicitly normalized because the isoperimetric number $i(G)$ itself is a normalized quantity in terms of number of vertices.

# Features Used in Our Experiments

Feature 1: $(p(G) - m_G(1))/|V(G)|$ as a lower bound of the number of pendant neighbors $q(G)$ normalized by $n = |V(G)|$ ;

Feature 2: The normalized Wiener index $W(G)/|V(G)|$ ;

Feature 3: $m_G(4,\infty)/|V(G)|$, i.e., the number of eigenvalues of $L(G)$ larger than $4$ (normalized) ;

Feature 4: $\sqrt{a(G)\left(2\max_{v\in V(G)} d_v - a(G)\right)}$, i.e., the upper bound of the isoperimetric number $i(G)$.

- We normalized Features 1, 2, 3, by $n = |V(G)|$ because we wanted to make features less dependent on the number of samples or how the dendrite arbors are sampled. Of course, the number of vertices itself could be a feature although it may not be a decisive one.
- Feature 4 was not explicitly normalized because the isoperimetric number $i(G)$ itself is a normalized quantity in terms of number of vertices.

# Features Used in Our Experiments

Feature 1: $(p(G) - m_G(1))/|V(G)|$ as a lower bound of the number of pendant neighbors $q(G)$ normalized by $n = |V(G)|$ ;

Feature 2: The normalized Wiener index $W(G)/|V(G)|$ ;

Feature 3: $m_G(4,\infty)/|V(G)|$, i.e., the number of eigenvalues of $L(G)$ larger than $4$ (normalized) ;

Feature 4: $\sqrt{a(G)\left(2\max_{v\in V(G)} d_v - a(G)\right)}$, i.e., the upper bound of the isoperimetric number $i(G)$.

- We normalized Features 1, 2, 3, by $n = |V(G)|$ because we wanted to make features less dependent on the number of samples or how the dendrite arbors are sampled. Of course, the number of vertices itself could be a feature although it may not be a decisive one.
- Feature 4 was not explicitly normalized because the isoperimetric number $i(G)$ itself is a normalized quantity in terms of number of vertices.

# Features Used in Our Experiments

Feature 1: $(p(G) - m_G(1))/|V(G)|$ as a lower bound of the number of pendant neighbors $q(G)$ normalized by $n = |V(G)|$ ;

Feature 2: The normalized Wiener index $W(G)/|V(G)|$ ;

Feature 3: $m_G(4, \infty)/|V(G)|$, i.e., the number of eigenvalues of $L(G)$ larger than $4$ (normalized) ;

Feature 4: $\sqrt{a(G)\left(2\max_{v \in V(G)} d_v - a(G)\right)}$, i.e., the upper bound of the isoperimetric number $i(G)$.

- We normalized Features 1, 2, 3, by $n = |V(G)|$ because we wanted to make features less dependent on the number of samples or how the dendrite arbors are sampled. Of course, the number of vertices itself could be a feature although it may not be a decisive one.
- Feature 4 was not explicitly normalized because the isoperimetric number $i(G)$ itself is a normalized quantity in terms of number of vertices.

# Features Used in Our Experiments ...

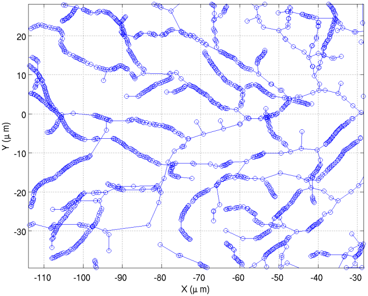
- *Feature 1* was used because the number of pendant neighbors seems to be strongly related to the so-called *spines*, short protrusions from the dendrite arbors.
- Hence, we expect that the larger this lower bound $p(G) - m_G(1)$ is, the more likely for the RGC to have spines.

# Features Used in Our Experiments . . .

- *Feature 1* was used because the number of pendant neighbors seems to be strongly related to the so-called *spines*, short protrusions from the dendrite arbors.
- Hence, we expect that the larger this lower bound $p(G) - m_G(1)$ is, the more likely for the RGC to have spines.

# Features Used in Our Experiments . . .

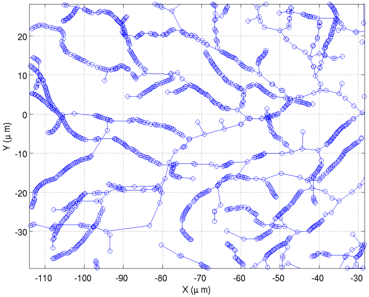
- *Feature 1* was used because the number of pendant neighbors seems to be strongly related to the so-called *spines*, short protrusions from the dendrite arbors.
- Hence, we expect that the larger this lower bound $p(G) - m_G(1)$ is, the more likely for the RGC to have spines.



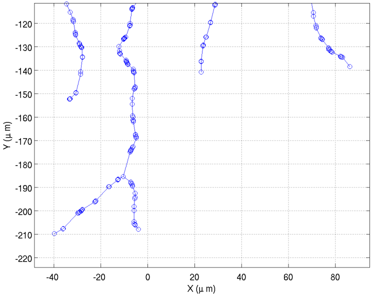

(a) RGC #60; $F_1$ large

# Features Used in Our Experiments . . .

- *Feature 1* was used because the number of pendant neighbors seems to be strongly related to the so-called *spines*, short protrusions from the dendrite arbors.
- Hence, we expect that the larger this lower bound $p(G) - m_G(1)$ is, the more likely for the RGC to have spines.




(a) RGC #60; $F_1$ large



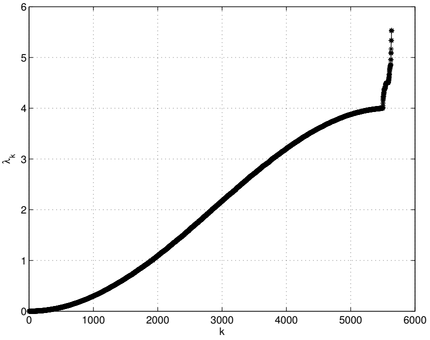

(b) RGC #100; $F_1$ small

# Features Used in Our Experiments . . .

- *Feature 3*, the normalized version of $m_G(4,\infty)$, was used because of the following observation:
- The eigenvalue distribution of each RGC consists of a smooth bell-shaped curve that ranges over $[0,4]$ and the sudden burst above the value 4.

# Features Used in Our Experiments . . .

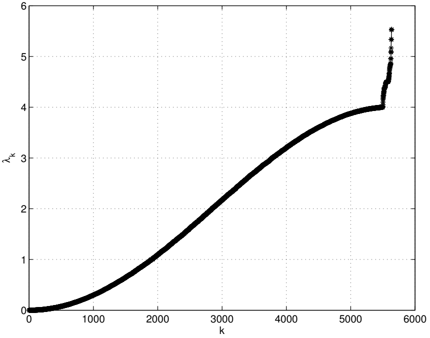
- *Feature 3*, the normalized version of $m_G(4,\infty)$, was used because of the following observation:
- The eigenvalue distribution of each RGC consists of a smooth bell-shaped curve that ranges over $[0,4]$ and the sudden burst above the value $4$.



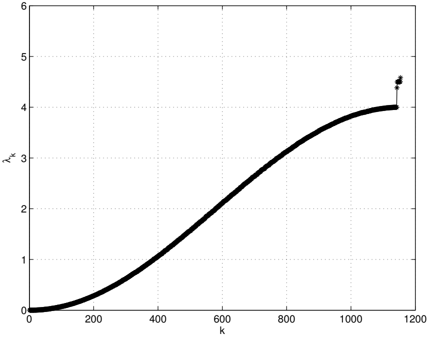

(a) RGC #60

# Features Used in Our Experiments . . .

- *Feature 3*, the normalized version of $m_G(4,\infty)$, was used because of the following observation:
- The eigenvalue distribution of each RGC consists of a smooth bell-shaped curve that ranges over $[0,4]$ and the sudden burst above the value $4$.




(a) RGC #60




(b) RGC #100

# Features Used in Our Experiments . . .

We have observed that this value 4 is critical since:

- the eigenfunctions corresponding to the eigenvalues below 4 are *semi-global* oscillations (like Fourier cosines/sines) over the entire dendrites or one of the dendrite arbors;
- those corresponding to the eigenvalues above 4 are much more *localized* (like wavelets) around branches.

# Features Used in Our Experiments . . .

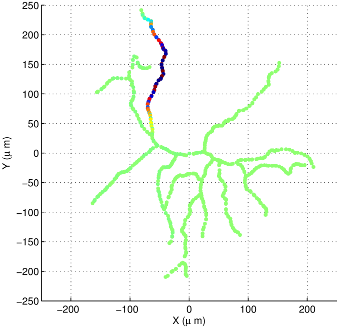
We have observed that this value $4$ is critical since:

- the eigenfunctions corresponding to the eigenvalues below $4$ are *semi-global* oscillations (like Fourier cosines/sines) over the entire dendrites or one of the dendrite arbors;
- those corresponding to the eigenvalues above $4$ are much more *localized* (like wavelets) around branches.

# Features Used in Our Experiments . . .

We have observed that this value 4 is critical since:

- the eigenfunctions corresponding to the eigenvalues below 4 are *semi-global* oscillations (like Fourier cosines/sines) over the entire dendrites or one of the dendrite arbors;
- those corresponding to the eigenvalues above 4 are much more *localized* (like wavelets) around branches.

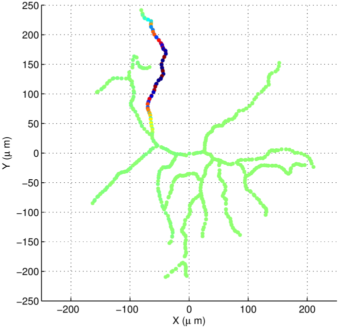



(a) RGC #100; $\lambda_{1141} = 3.9994$

# Features Used in Our Experiments . . .

We have observed that this value 4 is critical since:

- the eigenfunctions corresponding to the eigenvalues below 4 are *semi-global* oscillations (like Fourier cosines/sines) over the entire dendrites or one of the dendrite arbors;
- those corresponding to the eigenvalues above 4 are much more *localized* (like wavelets) around branches.

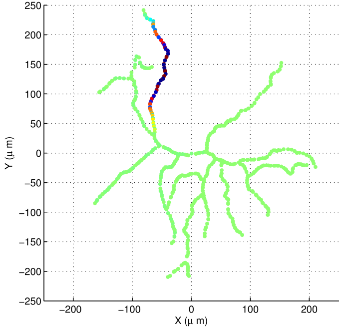



(a) RGC #100; $\lambda_{1141} = 3.9994$

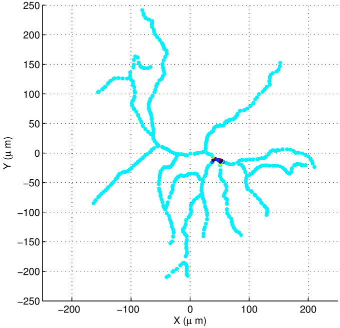
# Features Used in Our Experiments . . .

We have observed that this value 4 is critical since:

- the eigenfunctions corresponding to the eigenvalues below 4 are *semi-global* oscillations (like Fourier cosines/sines) over the entire dendrites or one of the dendrite arbors;
- those corresponding to the eigenvalues above 4 are much more *localized* (like wavelets) around branches.

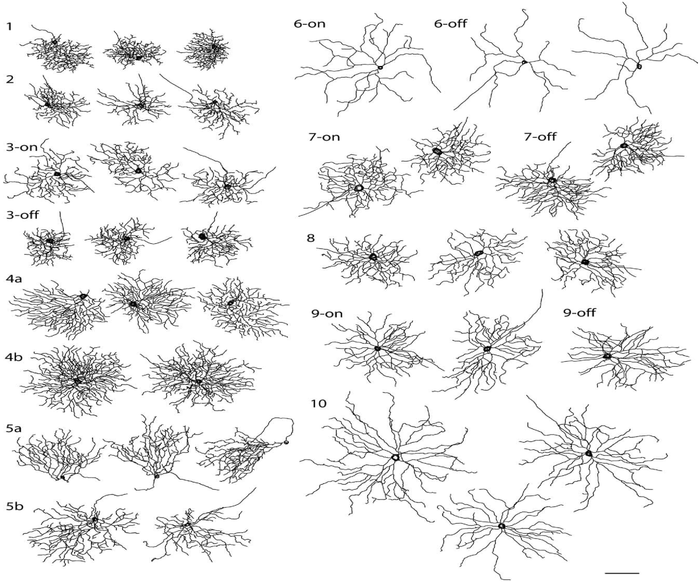



(a) RGC #100; $\lambda_{1141} = 3.9994$




(b) RGC #100; $\lambda_{1142} = 4.3829$

# Recap: Clustering using Features Derived by Neurolucida®



# Results: Scatter Plot; Feature 1 vs Feature 2

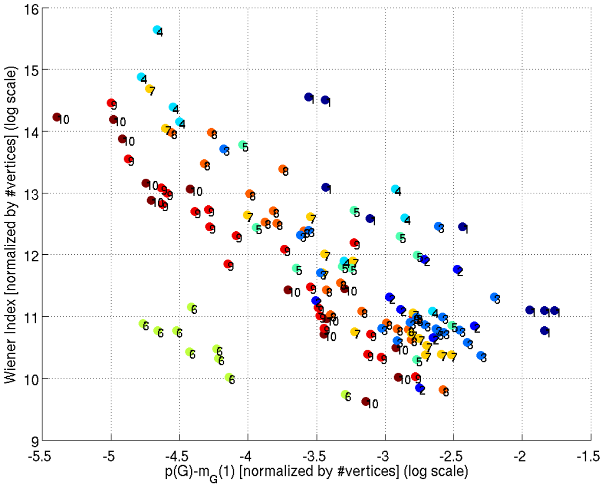



Figure: A scatter plot of the normalized lower bounds of the number of the pendant neighbors vs the normalized Wiener indices.

# Results: Scatter Plot; Feature 3 vs Feature 4

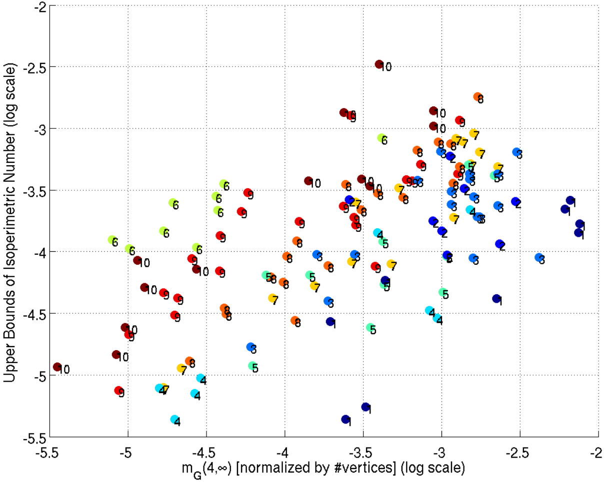
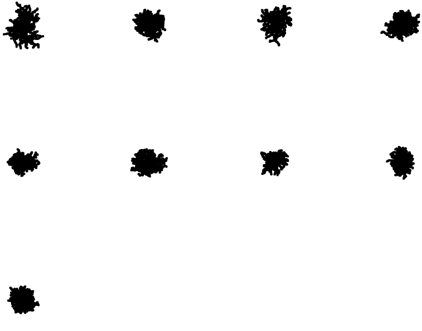



Figure: A scatter plot of the normalized number of the eigenvalues larger than $4$ vs the upper bounds of the isoperimetric numbers.

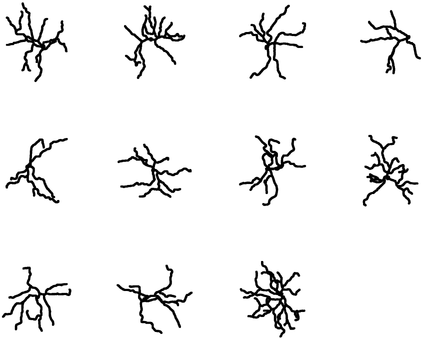
# Interpretation of the Results

- Cluster 6 RGCs separate themselves quite well from the other RGC clusters.
- In fact, the sparse and distributed dendrite patterns such as those in Clusters 6 and 10 are located below the major axis of the point clouds in the $F_1 - F_2$ scatter plot and above the major axis of the point clouds in the $F_3 - F_4$ scatter plot. $\Longrightarrow$ the dendrite patterns belonging to Cluster 6 and 10 have smaller number of spines and smaller Wiener indices compared to the other denser dendrite patterns such as Clusters 1 to 5.
- Considerable feature variability in Clusters 7 and 8.

# Cluster 1 vs Cluster 6 . . .



(a) Cluster 1



(b) Cluster 6

# Outline

# Conclusions & Future Plans

- Network-like structures are abundant and need to be quantitatively analyzed.
- How to embed such graphs/networks into a vector space becomes important.
- Demonstrated the usefulness of the eigenvalues of graph Laplacians for dendrite pattern analysis although the results are still preliminary.
- Need to investigate more eigenvalue-based features.
- Need to investigate resampling of dendrite arbor samples.
- How about the weighted graph Laplacians?
- Analyze the features derived by Neurolucida®: are they derivable from the Laplacian eigenvalues?
- Automating segmentation of dendritic trees from 3D images will be highly useful although it is quite tough.