Graph and Network Theoretic Modeling

ECS289I001, Fall 2016,
Lecture Notes.

Contents

Chapter 1

Introduction

1.1 Elements of a Graph

A graph, denoted by $G(V, E, W)$, is a combinatorial object with the following elements:

1. $V(G)$: the set of vertices/nodes.
2. $E(G)$: the set of links/edges.
3. $W(G) = \{w_{ij} \in \Re^+ : e_{ij} \in E(G)\}$ is the set of weights assigned to each edge of the graph[1] .

The edge weights are a measure of pairwise affinities (or separation) between adjacent nodes that, depending upon the context, may represent physical quantities such as link capacity in traditional communication networks; or acquaintance, amity or reciprocity in the social counterparts.

1.1.1 Degree and Volume

We denote by $n = |V(G)|$ the number of nodes in G, also called the *order* of the graph G. Similarly, $m = |E(G)|$ denotes the number of links. For $1 \leq i \leq n$, we define:

$$d(i) = \sum_{e_{ij} \in E(G)} w_{ij} \tag{1.1}$$

[1] $\Re^+$: the set of non-negative real numbers.

as the generalized degree of node i. Clearly, $d(i)$ is the aggregate affinity of node i with its one hop neighbors. The sum of node degrees in G, given as $Vol(G) = \sum_{i \in V(G)} d(i)$, is often referred to as the *volume* of the graph. $Vol(G)$ is, therefore, the aggregate one hop affinity in G. Without loss of generality, in the rest of this work, we deem the network represented by G to be connected. Or, in graph theoretic terms, G has exactly one connected component and, thus, at least one simple path between any pair of nodes.

1.1.2 Paths, Cycles and Walks

A *path* between a pair of nodes $(v_i, v_j) \in V(G) \times V(G)$ is a sequence of non-repeating edges starting at v_i and ending at v_j. Of all possible paths connecting a pair of nodes $(v_i, v_j) \in V(G) \times V(G)$, we are mostly interested in paths of shortest lengths. Henceforth, we denote by $SPD(i, j)$, the shortest path distance, between vertices v_i and v_j. Algorithmically, shortest paths can be computed using *Dijkstra's* algorithm, *Bellman-Ford* algorithm or *Floyd-Warshall* algorithm. A *cycle*, also referred to as a *simple cycle*, is a closed path starting and ending at the same vertex.

A *walk* from a node v_i to another node v_j, may traverse one or more edges multiple times. The analogue of a simple cycle, is a closed walk which by definition, is a walk that starts and ends at the same vertex.

1.1.3 Trees and Forests

A tree of order n, denoted by $T(n)$, is a simple, connected and acyclic graph with exactly $n - 1$ edges. Of particular interest are the *spanning trees* of a graph defined as:

Definition 1 *Spanning Tree of G: An acyclic subgraph of a graph G that connects all the vertices in G.*

For a weighted graph, a *minimum spanning tree*, denoted henceforth as $MST(G) =$, is defined as the spanning tree with the least sum of edge weights. For any given graph G, there can be more than one spanning and minimum spanning trees. The minimum spanning trees can be computed using either *Prim's* or *Kruskal's* algorithms.

A forest of order n, is simply a collection of trees. In particular, a forest of k edges, contains exactly $n - k$ trees in it. All analogues of trees - spanning and minimum spanning - also apply to forests. Spanning trees and forests are indicative of the path

diversity in a graph and are therefore used in multiple graph analysis problems. We shall return to these later.

1.2 Algebraic Representation of a Graph

1.2.1 The Adjacency Matrix

The adjacency matrix of $G(V, E, W)$ is defined as $\mathbf{A} \in \Re^{n \times n}$, with elements $[\mathbf{A}]_{ij} = a_{ij} = a_{ji} = [\mathbf{A}]_{ji} = w_{ij}$, if $i \neq j$ and $e_{ij} \in E(G)$ is an edge; 0 otherwise. As stated earlier, the value w_{ij} is a measure of one hop *affinity* between nodes i and j. By definition, $\forall 1 \leq i \leq n : \quad A_{ii} = 0$. Clearly, $\mathbf{A}$ is a real and symmetric matrix.

The degree matrix of $G(V, E, W)$, is a diagonal matrix $\mathbf{D} \in \Re^{n \times n}$ such that $[\mathbf{D}]_{ii} = d_{ii} = d(i) = \sum_{j \in V(G)} a_{ij}$, is the weighted degree of node $i \in V(G)$.

1.2.2 The Laplacian and it's Pseudo-Inverse

The combinatorial Laplacian of the graph is then given by:

$$\mathbf{L} = \mathbf{D} - \mathbf{A} \tag{1.2}$$

Despite its simple form, the structure and eigen spectrum of $\mathbf{L}$ account for significant topological characteristics of the graph, such as minimal cuts, clustering and determining the number of spanning trees [1, 2, 3, 4]. The Laplacian thus finds use in various aspects of structural analysis [5, 6, 7, 2, 3, 8, 9, 10, 11, 12, 13, 14, 4]. It is easy to see, from the definition in (1.2) above, that the Laplacian $\mathbf{L}$ is a real, symmetric and doubly-centered matrix:

$$\sum_{i=1}^{n} [\mathbf{L}]_{ij} = \sum_{j=1}^{n} [\mathbf{L}]_{ij} = 0 \tag{1.3}$$

More importantly, $\mathbf{L}$ admits an eigen decomposition of the form:

$$\mathbf{L} = \mathbf{\Phi} \mathbf{\Lambda} \mathbf{\Phi}' \tag{1.4}$$

where the columns of $\mathbf{\Phi} = [\phi_1, \phi_2, ..., \phi_{n-1}, \phi_n]$ constitute the set of eigen vectors of $\mathbf{L}$. For the combinatorial Laplacian $\mathbf{L}$, this set is orthogonal [15]:

$$\phi_1 \perp \phi_2 \perp ... \perp \phi_{n-1} \perp \phi_n : \qquad \phi_i \cdot \phi_j = 0, \qquad \forall 1 \leq i \neq j \leq n \tag{1.5}$$

where $(\cdot)$ is the inner/dot product operator for vectors. Therefore, $\mathbf{\Phi}$, with appropriate normalization, constitutes the orthonormal basis of $\Re^n$. Also, $\mathbf{\Lambda}$ is a diagonal matrix with $[\Lambda]_{ii} = \lambda_i : 1 \leq i \leq n$; being the n eigen values of $\mathbf{L}$. It is well established that for an undirected graph $G(V, E, W)$, $\mathbf{L}$ is positive semi-definite i.e. all its eigen values are non-negative [15]. Further, if G is connected, as we have assumed, the smallest eigen value of 0 is unique. By convention, we shall assume a descending order for the eigen values of $\mathbf{L}$:

$$\mathbf{\Lambda} = [\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_{n-1} > \lambda_n = 0] \tag{1.6}$$

Note $\lambda_n = 0$, implies that $\mathbf{L}$ is rank deficient ($rank(\mathbf{L}) = n-1 < n$) and consequently singular. Its inverse, in the usual sense, does not exist.

However, a generalized inverse, called the Moore-Penrose pseudo-inverse of $\mathbf{L}$, denoted henceforth by $\mathbf{L}^+$, does exist and is unique [15]. Following four constitute the defining properties of $\mathbf{L}^+$:

$$a.\ \ \mathbf{L}\mathbf{L}^+\mathbf{L} = \mathbf{L} \qquad b.\ \ \mathbf{L}^+\mathbf{L}\mathbf{L}^+ = \mathbf{L}^+ c.\ \ (\mathbf{L}\mathbf{L}^+)' = \mathbf{L}\mathbf{L}^+ \qquad d.\ \ (\mathbf{L}^+\mathbf{L})' = \mathbf{L}^+\mathbf{L} \tag{1.7}$$

Like $\mathbf{L}$, $\mathbf{L}^+$ is also real, symmetric, doubly centered and positive semi-definite. Moreover, the eigen decomposition of $\mathbf{L}^+$ is given by

$$\mathbf{L}^+ = \Phi\Lambda^+\Phi' \tag{1.8}$$

with the same set of orthogonal eigen-vectors as that of $\mathbf{L}$. The set of eigen values of $\mathbf{L}^+$, given by the diagonal matrix $\mathbf{\Lambda}^+$, is composed of $\lambda_n^+ = 0$ and the reciprocals of the positive eigen-values of $\mathbf{L}$, i.e. $[\lambda_1^{-1} \leq \lambda_2^{-1} \leq ... \leq \lambda_{n-1}^{-1}]$. It is the eigen space of $\mathbf{L}^+$, derived from the eigen space of $\mathbf{L}$, that is of interest to us.

Defining $\mathbf{X} = \mathbf{\Lambda}^{+1/2}\mathbf{\Phi}'$, we obtain:

$$\mathbf{L}^+ = \mathbf{\Phi}\mathbf{\Lambda}^+\Phi' = \mathbf{X}'\mathbf{X} \tag{1.9}$$

The form in (1.9) above, together with the fact that Φ is an orthonormal basis for $\Re^n$, implies that the matrix $\mathbf{X}$ represents an embedding of the network in an $n-$ dimensional Euclidean space (cf. [8, 16] and the references therein). Next, we describe the specifics of this embedding space, with respect to the nodes and edges in the graph.

Let $\mathbf{x}_i$ denote the i^{th} column of $\mathbf{X}$. For a node $i \in V(G)$, $\mathbf{x}_i$ represents an n-dimensional co-ordinate in the embedding. In other words, $\mathbf{x}_i$ is the position vector

for the node i in this n-dimensional space. Also, as $\mathbf{L}^+$ is doubly-centered (rows and columns individually sum up to 0), the centroid of set of all node position vectors, lies at the origin of the n-dimensional space. Thus, the squared distance of node i from the origin (or the squared length of the position vector) corresponds to the i^{th} diagonal entry in $\mathbf{L}^+$:

$$||\mathbf{x}_i||_2^2 = [\mathbf{L}^+]_{ii} = l_{ii}^+ \tag{1.10}$$

On the other hand, for any pair of nodes $(i, j) \in V(G) \times V(G)$, the embedding yields two geometric quantities. First, the inner product of the position vectors:

$$\mathbf{x}_i \cdot \mathbf{x}_j = [\mathbf{L}^+]_{ij} = l_{ij}^+ = l_{ji}^+ = [\mathbf{L}^+]_{ji} = \mathbf{x}_j \cdot \mathbf{x}_i \tag{1.11}$$

And secondly, the pairwise distance between nodes:

$$||\mathbf{x}_i - \mathbf{x}_j||_2^2 = l_{ii}^+ + l_{jj}^+ - l_{ji}^+ + l_{ij}^+ = ||\mathbf{x}_j - \mathbf{x}_i||_2^2 \tag{1.12}$$

Finally, we conclude by defining the volume of the overall embedding as follows:

$$\Xi(\mathbf{L}^+) = \sum_{i=1}^{n} l_{ii}^+ = Tr(\mathbf{L}^+) \tag{1.13}$$

where $Tr(\mathbf{L}^+)$ is the *trace* of the matrix $\mathbf{L}^+$. The volume of the embedding gives us a measure of *compactness* and is an aggregate quantity defined for the network as a whole.

We therefore obtain a Euclidean embedding of the network in terms of the eigen space of $\mathbf{L}^+$ with each node in the network represented as a point and a distance defined over any arbitrary node pair. This is indeed the central construct upon which this thesis relies. In particular, note that each of the geometric attributes described above (except $\mathbf{x}_i \cdot \mathbf{x}_j$), is determined by the metric distance defined for the Euclidean space. It is this function that holds the key to our analyses in the subsequent chapters; and it has an intriguing analogue in the physical world too as we explain next.

1.3 Equivalent Electrical Network, Effective Resistance Distance and $\mathbf{L}^+$

An interesting analogy exists between simple undirected graphs and resistive electrical circuits [17, 18]. Given a simple, connected and undirected graph $G(V, E, W)$, the

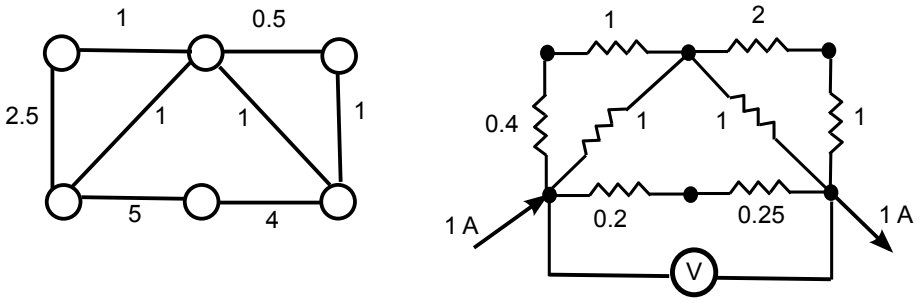

Figure 1.1: A simple graph G and its EEN.

equivalent electrical network (EEN) of the graph can be formed by replacing each edge $e_{ij} \in E(G)$, of affinity w_{ij}, with an electrical resistance $\omega_{ij} = w_{ij}^{-1}$ *ohm* (cf. Fig. 1.1). Needless to say, each node $i \in V(G)$ has an analogous *junction* point in the EEN. A distance function can then be defined between any pair of nodes $(x, y) \in V(G) \times V(G)$ in the resulting EEN as follows:

Definition 2 *Effective Resistance (Ω_{ij}): The voltage developed between nodes i and j, when a unit current (1 amp) is injected at node i and is withdrawn at node j.*

It is well established that the square root of the effective resistance distance ($\sqrt{\Omega_{ij}}$) is a Euclidean metric [8, 18], i.e. it satisfies the following properties:

a. **Identity**: $\sqrt{\Omega_{ij}} > 0$ if $i \neq j$, 0 otherwise.

b. **Symmetry**: $\sqrt{\Omega_{ij}} = \sqrt{\Omega_{ji}}$

c. **Triangle-inequality**: $\sqrt{\Omega_{ij}} \leq \sqrt{\Omega_{ik}} + \sqrt{\Omega_{jk}}$

Most importantly, Ω_{ij} can be expressed in terms of the elements of $\mathbf{L}^+$ as follows [8]:

$$\Omega_{ij} = l_{ii}^+ + l_{jj}^+ - l_{ij}^+ - l_{ji}^+ \tag{1.14}$$

Therefore, the distance metric that our Euclidean space is endowed with is simply the effective resistance distance — an equivalence that we exploit time and again with significant rewards.

1.4 Summary

In this lecture, we introduced the preliminary notations and algebraic paraphernalia that are used throughout this thesis. We described an n-dimensional Euclidean embedding of the network in terms of the eigen space of the Moore-Penrose pseudo-inverse of the combinatorial Laplacian and established how the metric distance function characterizing this space is indeed the same as the pairwise effective resistance distance, a physical analogue in the real world. In subsequent chapters, we use these geometric notions to characterize the robustness of a network at all granularities — from individual nodes and edges to the network as a whole.

References

[1] N. Biggs. *Algebraic Graph Theory*. Cambridge University Press, 1993.

[2] M. Fiedler. Algebraic connectivity of graphs. *Czechoslovak Math. J.*, 23:298–305, 1973.

[3] M. Fiedler. A property of eigenvectors of nonnegative symmetric matrices and its applications to graph theory. *Czechoslovak Math. J.*, 25(100):619–633, 1975.

[4] U. V. Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416, December 2007. Max Planck Institute for Biological Cybernetics. Technical Report No. TR-149.

[5] Noga Alon. Eigenvalues and expanders. *Combinatorica*, 6(2):83–96, 1986.

[6] F. R. Chung. *Spectral Graph Theory*. Am. Math. Soc., 1997.

[7] W. E. Donath and A. J. Hoffman. Lower bounds for the partitioning of graphs. *IBM J. of Res. and Dev.*, 17:410–425, 1973.

[8] F. Fouss, A. Pirotte, J. M. Renders, and M. Saerens. Random-walk computation of similarities between nodes of a graph with application to collaborative recommendation. *IEEE Transactions on Knowledge and Data Engineering*, 19, 2007.

[9] F. Gobel and A. Jagers. Random walks on graphs. *Stochastic Processes and Their Applications*, 2:311–336, 1974.

[10] B. Mohar. The Laplace spectrum of graphs. In *Graph Theory, Combinatorics, and Applications*, volume 2, pages 871–898. Wiley, 1991.

[11] B. Mohar. Laplace eigenvalues of graphs - a survey. *Discrete Math.*, 109:171–183, 1992.

[12] B. Mohar. Some applications of Laplace eigenvalues of graphs. In *Graph Symmetry: Algebraic Methods and Applications*, volume NATO ASI Ser C 497, pages 225–275. Kluwer, 1997.

[13] A. Nilli. On the second eigenvalue of a graph. *Discrete Mathematics*, 91(2):207–210, 1991.

[14] J. Shi and J. Malik. Normalised cuts and image segmentation. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 22(8):888–905, 1997.

[15] A. Ben-Israel and T. Greville. *Generalized Inverses: Theory and Applications,* 2^{nd} *edition.* Springer-Verlag, 2003.

[16] M. Saerens, F. Fouss, L. Yen, and P. Dupont. The principal components analysis of a graph, and its relationships to spectral clustering. In *Proc. 15th European Conf. Machine Learning (ECML '04)*, pages 371–383, 2004.

[17] P. G. Doyle and J. L. Snell. *Random Walks and Electric Networks.* The Math. Assoc. of America, 1984.

[18] D. J. Klein and M. Randić. Resistance distance. *J. Math. Chemistry*, 12:81–95, 1993.