Social and Technological Networks	Edinburgh, 2018

Notes 3. Distances and metric spaces.

Rik Sarkar *Notes*

Measurements of distances form the core of all analysis, including machine learning and network analysis. Given a set X, the distance between elements $u, v \in X$ is defined by a function $d : X \to \mathbb{R}$. So, $d(u, v)$ returns a real number giving the distance between u and v.

The familiar distance between points in a plane is called the Euclidean distance.

Q 1. *What is the function expressing Euclidean distance between points in a D dimensional space?*

Sometimes, the Euclidean distance is called the L_2 norm. And the distance is written as $|u - v|$ or $||u - v||$.

The L_1 norm is sometimes written as $|u - v|_1$. In $2 - D$ plane, the L_1 norm is $|u_1 - v_1| + |u_2 - v_2|$. In general, the L_p distance in plane is defined as $|u - v|_p = \sqrt[p]{(u_1 - v_1)^p + (u_2 - v_2)^p}$.

(How does the definition change in higher dimensions?)

So, distances between points or items can be measured in different ways. In general, we a re interested in metric spaces:

Definition 0.1 (Metric Spaces). *A metric space consists of a set X and a distance function $d : X \to \mathbb{R}$, such that, for any $u, v, w \in X$, the following are true:*

- $d(u, v) \geq 0$
- $d(u, v) = 0$ *if and only if* $u = v$
- $d(u, v) = d(v, u)$
- $d(u, v) \leq d(u, w) + d(w, v)$

Check that these properties hold for points in a plane.

Now suppose that we measure distances on a graph using edge weights. let us take a complete graph G, with weight on any edge given by $w(u, v)$. We can treat the weight as a distance on the set V. That is, the distance $d(u, v)$ between any two vertices is given by $d(u, v) = w(u, v)$.

Q 2. *Take the complete graph on some fixed number of vertices, say 4 or 5. In what ways can you define w such that it is not a metric? [hint: create edge weights such that (one or some of them) violate a condition of being a metric. Repeat for the other metric conditions.]*

Shortest path distance is a distance measure where the $d(u, v)$ is defined as the total weight on the shortest (least weight) path between u and v.

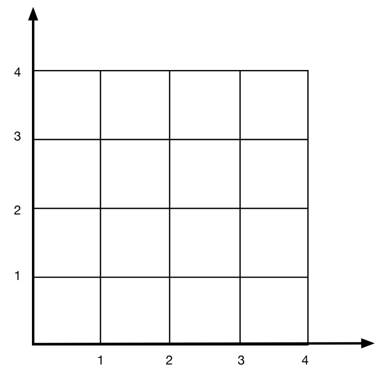

Figure 1. Grid graph embedded in Euclidean plane.

Q 3. *Suppose in a weighted, undirected, connected graph, all the weights are positive (greater than zero). Show that shortest path distance is a metric. (note that this holds even when the graph is not complete.)*

Intrinsic and extrinsic metrics. We can think of multiple types of distances between vertices in a network. For example, we can embed the vertices of a graph into a plane. Say, $f : V \to \mathbb{R}^2$ is function that assigns coordinates to each vertex. This gives us at least two ways to measure distances between u and v:

1. **Extrinsic metric**. The Euclidean distance between the assigned coordinates of vertices: $|f(u)- f(v)|$.
2. **Intrinsic metric**. The shortest path distance from u to v, where the weight of each edge ab is taken as the distance between their Euclidean length in the embedding: $|f(a) - f(b)|$.

If the graph had previously assigned weights w, then the distance defined by w (let us call it d_w) is yet another intrinsic metric.

Both intrinsic and extrinsic metric can have different forms, and we can define multiple of each. An intrinsic metric can be realised by walking along the edges present in the graph. So we can go from u to v by traveling the intrinsic distance between them, by sticking to the edges of the graph. While to get from u to v by traveling what the extrinsic metric gives, we have to walk outside the graph along the space where it is embedded.

Q 4. *What are the intrinsic and extrinsic metrics for the grid graph embedded in Euclidean plane?*