Metrics: Growth, dimension, expansion

Social and Technological Networks

Rik Sarkar

University of Edinburgh, 2017.

Metric

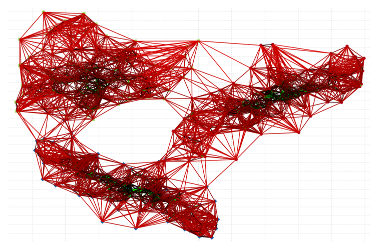
- A distance measure d is a metric if:
  - $d(u,v) \geq 0$
  - $d(u,v) = 0$ iff $u=v$
  - $d(u,v) = d(u,v)$
  - $d(u,v) \leq d(u,v) + d(u,v)$

Metrics

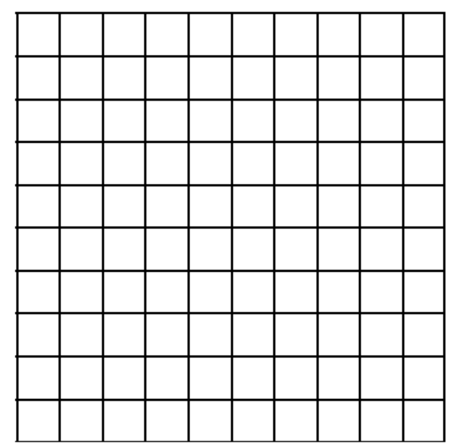
- Metrics are Important because:
  - Metrics are used to construct networks
  - Networks have metrics that determine their properties
- Today:
  - Finding metric distances in data
  - Properties of some typical metrics

Graph Embedding

- Map the vertices V to points in the plane
  - (or some other space)

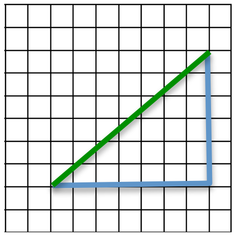


- Usually, different vertices are mapped to different points

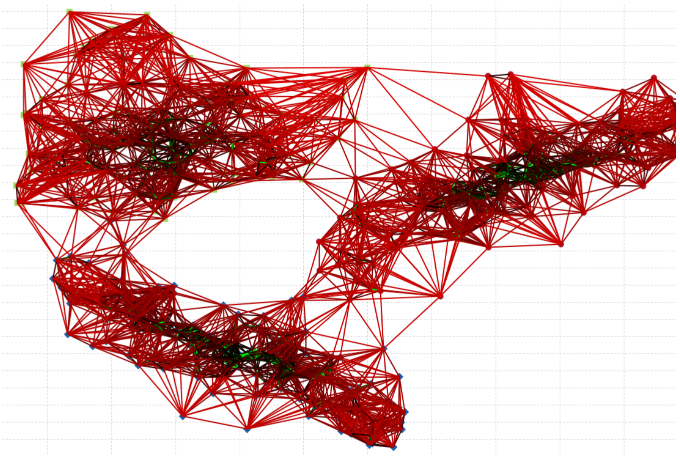


Different distances

- What is the distance between u and v?
- Possibility 1 (Embedding or extrinsic distance):
  - Distance in the embedded space
    - E.g. Euclidean distance



- Possibility 2 (Intrinsic distance):
  - Distance in the graph
    - The length of shortest path



- Possibility 3 (Intrinsic distance):
  - Weighted distance in the graph
    - Weight of least weight path

Where do metrics come from?

- Possibility 1:
  - Vertex locations are given. Eg. Mobile phone locations
- Possibility 2:
  - We are given real valued features like age, salary, etc
  - We can use these as dimensions and compute distances.

Computing distances for categorical data

- Suppose we are given categorical data
- E.g.
  - We are given list of clubs people belong to
  - Or list of songs they like etc..
- Cosine distance
  - Represent the list as 0-1 vectors A, B, …
  - Find cosine similarity $S_c = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|_2 \|\mathbf{B}\|_2}$
  - Cosine distance $d_c = 1 - S_c$

Computing distances for categorical data

- Jaccard similarity:
  - Treat the vectors as sets A, B..
  - And compute $$J(A,B) = \frac{|A \cap B|}{|A \cup B|}$$
  - Distance J_d = 1-J

Computing distances for categorical data

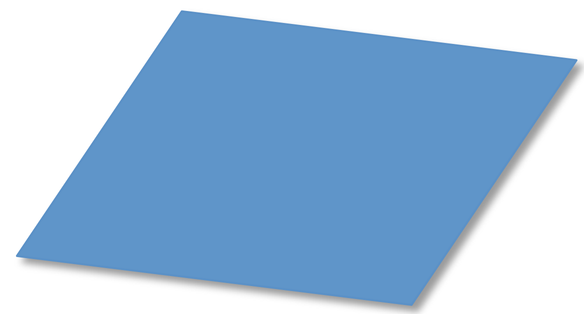
- Min category distance
  - Take the size of the smallest club with both A and B as the distance

- You can come up with many other ways of computing distance

Euclidean metric

- 1-D
  - Straight line (think x-axis)

- 2-D
  - Plane



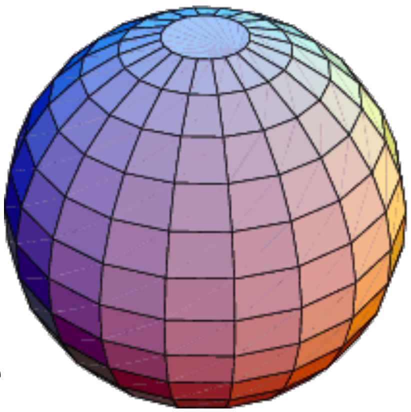
- Distance measure in dimension d:

$$d(u, v) = \sqrt{(u_1 - v_1)^2 + (u_2 - v_2)^2 + \cdots + (u_d - v_d)^2}$$

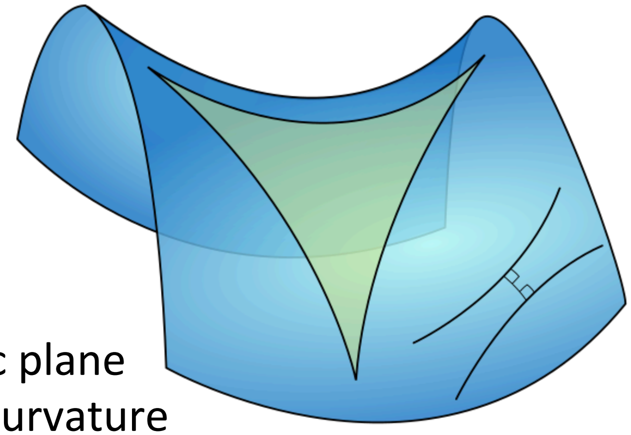
Non Euclidean metrics

- A lot of maths for Euclidean metrics
- What are examples of non-Euclidean metrics?

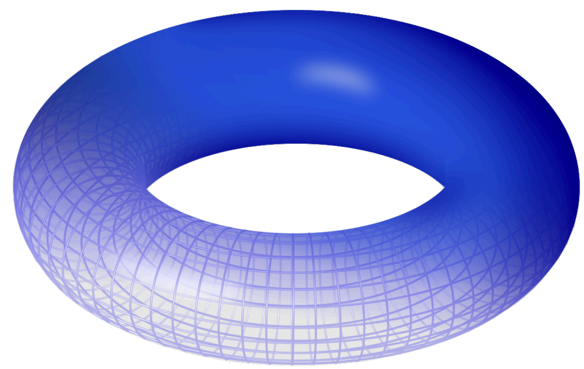
Non-euclidean metrics



Sphere
Positive curvature



Hyperbolic plane
Negative curvature



Realistic shapes. With bends, and cycles.

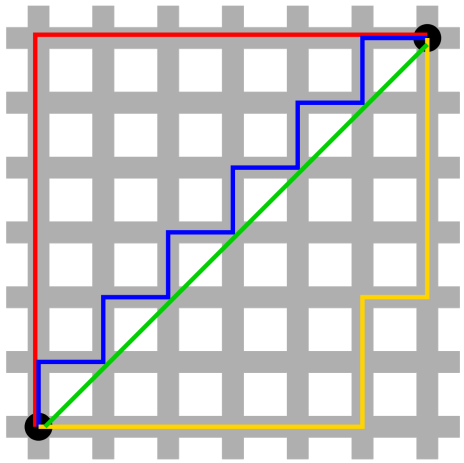
- L_p metrics

$$d(u,v) = \sqrt[p]{(u_1 - v_1)^p + (u_2 - v_2)^p + \cdots + (u_d - v_d)^p}$$

L_1 metric

- Manhattan distances

$$d(u, v) = |u_x - v_x| + |u_y - v_y|$$



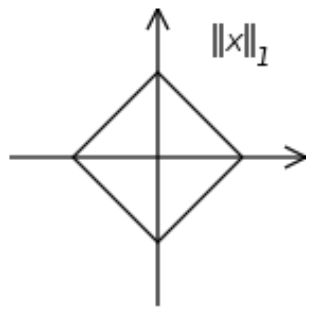
L_∞ Metric

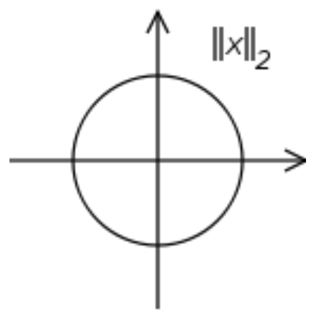
- Largest component over dimensions

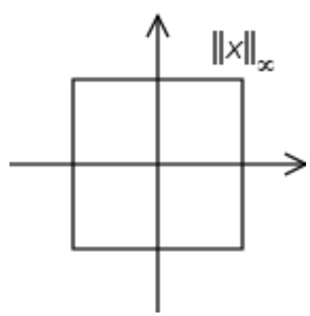
$$d(u,v) = \lim_{p \to \infty} \sqrt[p]{(u_x - v_x)^p + (u_y - v_y)^p}$$

$$d(u,v) = \max(|u_x - v_x|, |u_y - v_y|)$$

Disks and circles in L_p metrics

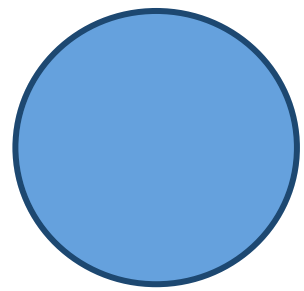






Ball

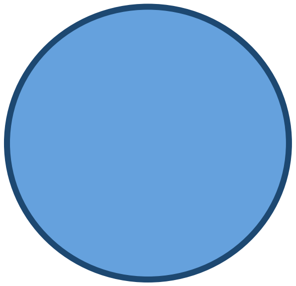
- A ball of radius r at point v:
  - The set of all points within distance r from v
  - Called a disk in 2D
- Usually written as
  - B(v,r) or
  - $B_r(v)$



- Sphere $S_r(v)$: set of points at distance exactly r from v
  - The boundary of the ball
  - 1-D sphere: boundary of a 2-D ball
  - 2-D sphere: boundary of a 3-D ball etc

Size of a ball

- The “measure” in a suitable dimension
  - Area in 2D
  - Volume in 3D etc
  - What about 1D?



- What is the measure of a sphere?
  - 1D?
  - 2D?

Growth of a metric

- How does the size of a ball B(v,r) grow with radius?
- In 1D?
- In 2D?
- In 3D?

Growth of Euclidean metric

- D-dim

$$\Theta(r^d)$$

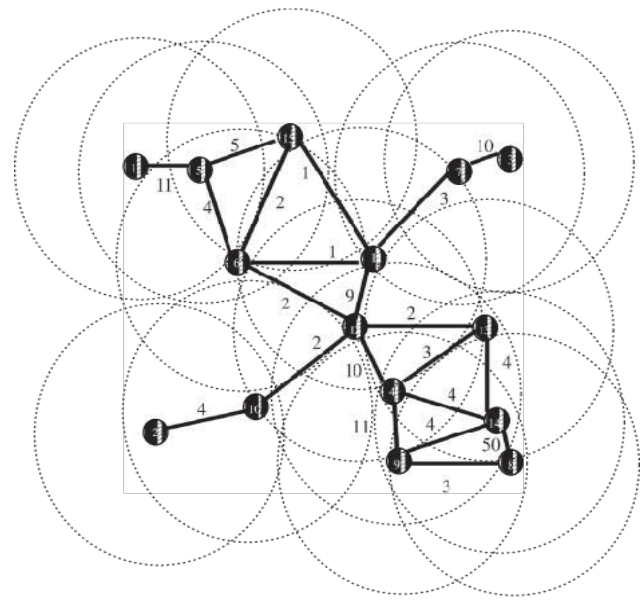
Growth can be used to detect dimension

- E.g. Long strips

  - Growth is linear: O(r)
  - Compared to size, this is 1-D

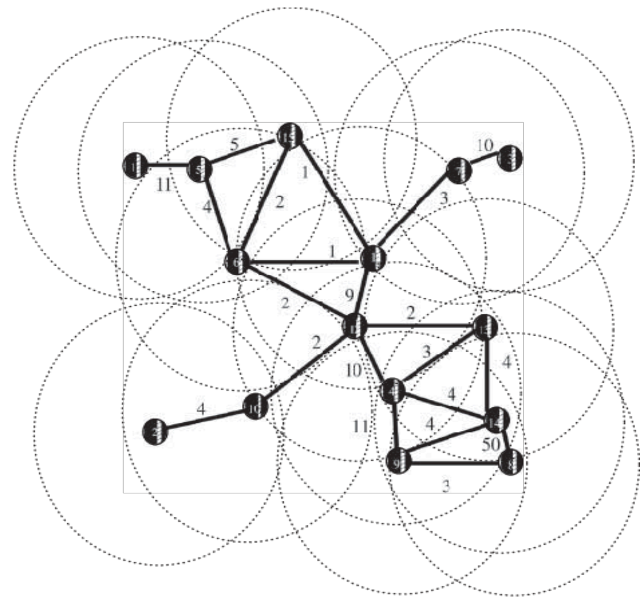
Making networks from metrics

- Unit disk graphs
  - Consider vertices in the plane (like wireless nodes)
  - Connect two vertices by an edge if they are within distance 1 of each-other. (within transmission distance)
  - Applies generally to higher dim (Unit ball graphs)
  - Connect two nodes if they are within a given distance



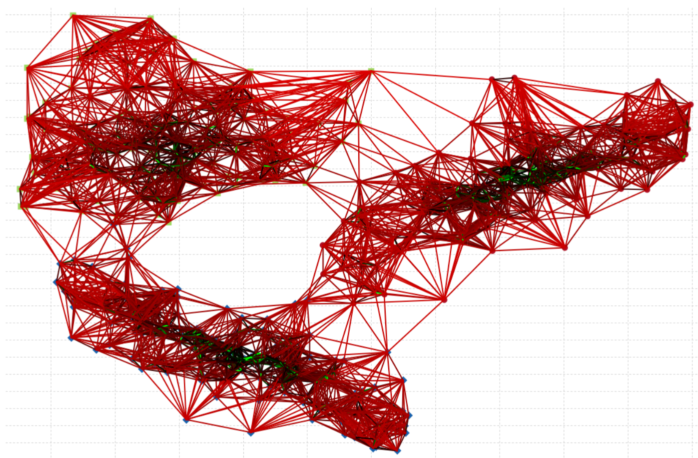
Making networks from metrics

- Unit disk graphs
  - Consider vertices in the plane (like wireless nodes)
  - Connect two vertices by an edge if they are within distance 1 of each-other. (within transmission distance)
  - Applies generally to higher dim (Unit ball graphs)
  - Connect two nodes if they are within a given distance



k-NN graphs

- For each vertex, fine k nearest neighbors
  - Connect edges to all k nearest neighbors
  - Variants:
    - Connect all k-NN edges
    - Connect only if both vertices are k-NN of each-other



The undirected graph distance

- Is a metric
- In unweighted graphs, all values are integers

Finding distance between two nodes in a graph

- Breadth first search
- Dijkstra's shortest path algorithm

Ball B(v,r) in a graph

- The set of nodes within distance r of v
  - Sphere: Nodes at distance exactly r
- Measure or volume:
  - The number of nodes in B(v, r)
- Growth
  - How does |B(v, r)| grow with r
  - For Chain?
  - Cycle?
  - Grid?
  - Balanced binary tree?

Growth in chain graph

- When the graph is a chain. What happens when we use
  - Extrinsic metric Euclidean L_2
  - Extrinsic metric L_1
  - Extrinsic metric L_∞
  - When we use the intrinsic graph metric

Growth in chain graph

- When the graph is a chain. What happens when we use
  - Extrinsic metric Euclidean L_2
  - Extrinsic metric L_1
  - Extrinsic metric L_∞
  - When we use the intrinsic graph metric

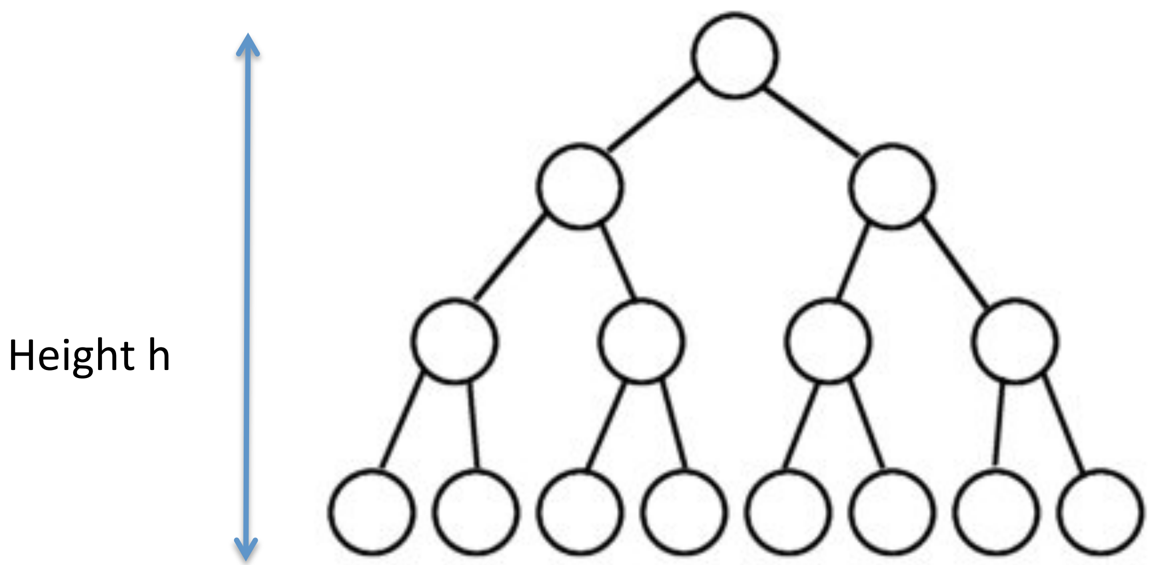
  - When the chain is embedded differently?

Growth in Cycle

Growth in 2D grid

Growth for balanced binary trees

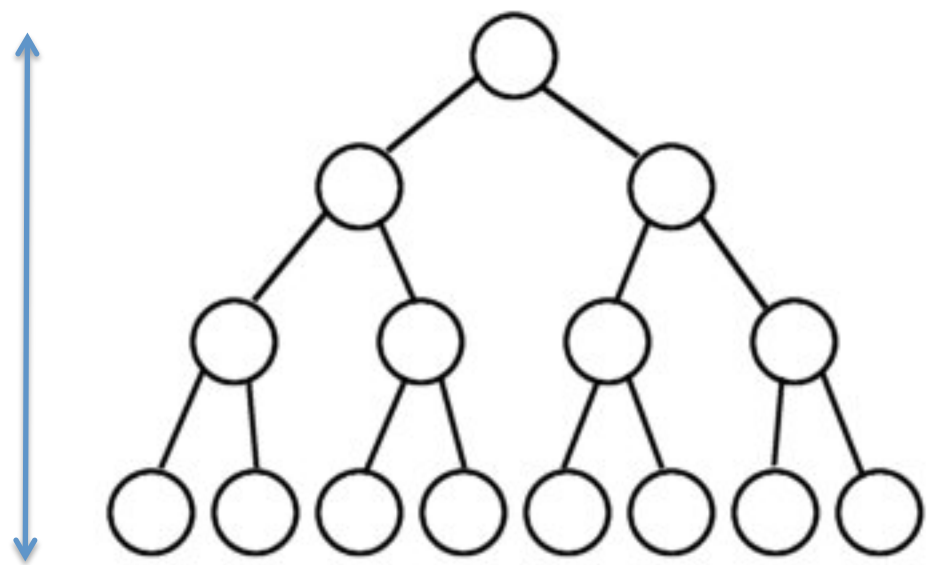
•



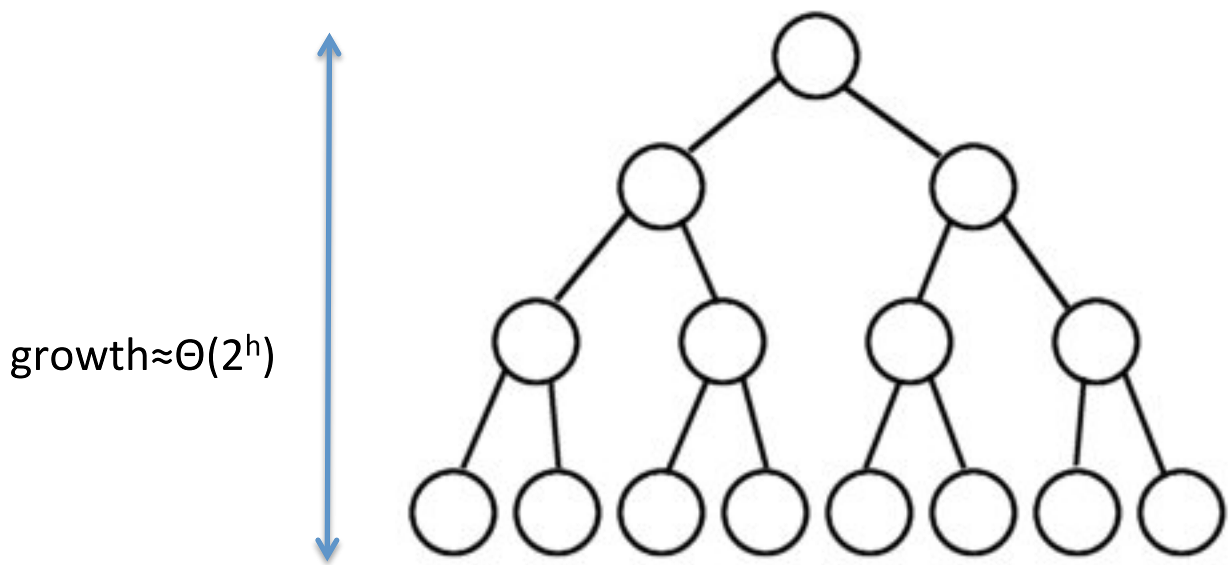
Growth for balanced binary trees

-

$n=2^h$



Growth for balanced binary trees



What are the diameters of these graphs

- Finite chain
- Finite Grid
- Balanced binary tree

- Intrinsic metric
- Extrinsic metric

How does the sphere grow?

- For chain
- Grid
- Balanced binary tree

- Intrinsic metric
- Extrinsic metric

Edge Expansion

- How fast the ‘boundary’ expands relative to ‘volume’ or ‘size’ of a subset
- Boundary of S :
  - $e^{out}(S)$: edges with exactly one end-point in S
- Expansion:

$$\alpha = \min_{S \subseteq V} \frac{|e^{out}(S)|}{\min(|S|, |\overline{S}|)}$$

Expansion

$$\alpha = \min_{S \subseteq V} \frac{|e^{out}(S)|}{\min(|S|, |\overline{S}|)}$$

- Equivalently:

$$\alpha = \min_{|S| \leq n/2} \frac{|e^{out}(S)|}{|S|}$$

Expanders

- A class of graphs with expansion at least a constant $\alpha \geq c$
  - For some constant c

Are the following graphs expanders?

- A chain
- A balanced binary tree
- A grid

Examples of expanders

- Random d-regular graphs for $d>3$

- ER graphs for large enough p

Expanders have small diameter

- A graph with degrees $\leq$ d and expansion $\geq \alpha$
- Has diameter

$$O(\tfrac{d}{\alpha} \lg n)$$

Other properties

- Expanders are well connected
- Usually sparse (number of edges much smaller than n^2)
- Diffusion processes spread fast in an expander
- Random walks mix fast (achieve steady state)
- …

- Do exercise 2.
- Ex 3 to be given out soon
- Study the material up to now. This will important later.

Project

- Suggested list to be announced Monday/Tuesday
  - On Email and Piazza
- 4 days to select project
  - You can work in groups of 1,2 or 3
  - But Must submit individual reports
  - The group is for discussion and possibly some common tasks. But marking is individual
- A short (half page to 1 page) proposal due around October 27.
  - Not graded. Feedback only
- Final report due around Nov 15.