Non-euclidian data and Graphs

Raoul Grouls, 10 januari 2024

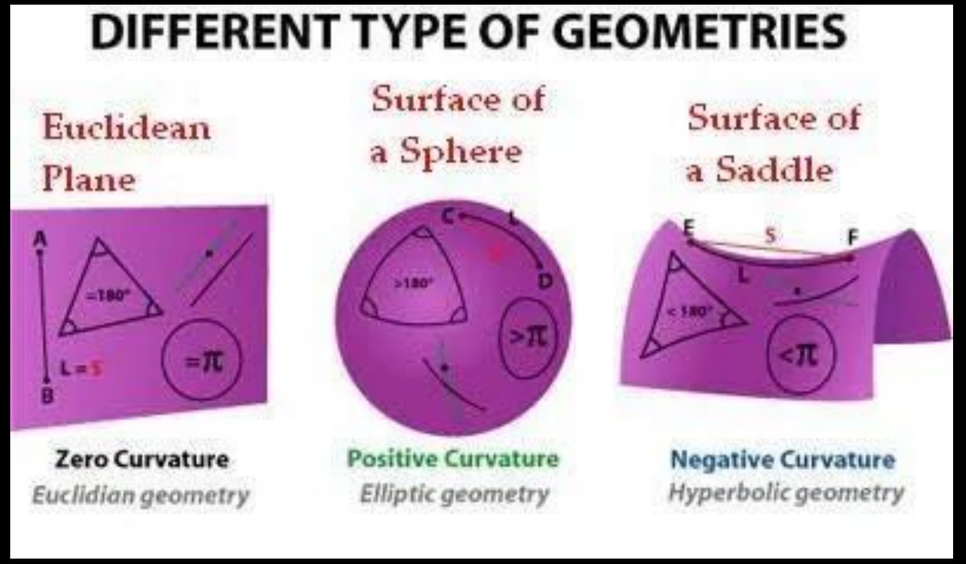
What is Euclidian geometry?

Euclidian data follows the axioms of euclidian geometry

1. A straight line may be drawn between any two points.
2. Any terminated straight line may be extended indefinitely.
3. A circle may be drawn with any given point as center and any given radius.
4. All right angles are equal.
5. For any given point not on a given line, there is exactly one line through the point that does not meet the given line

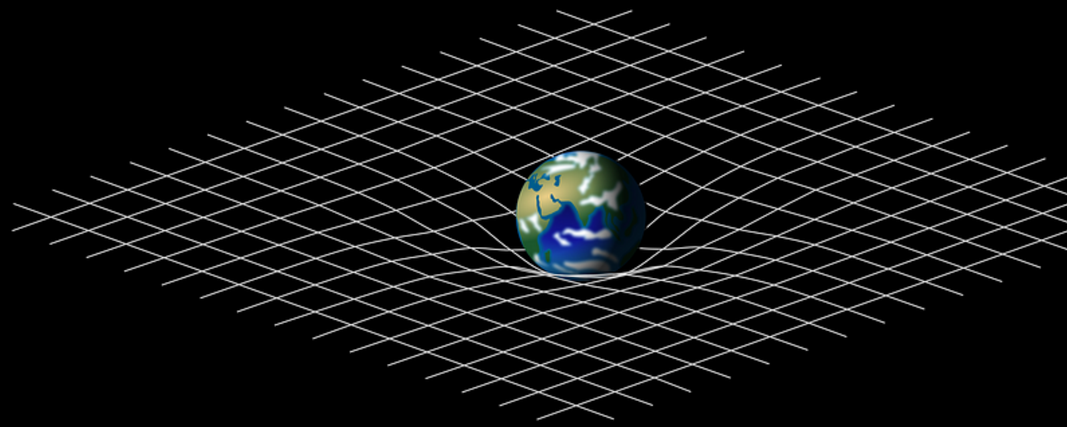
What is Non-Euclidian geometry?

Rejection of the parallel postulate



Early 19th century, the parallel postulate was rejected as “apriori true”

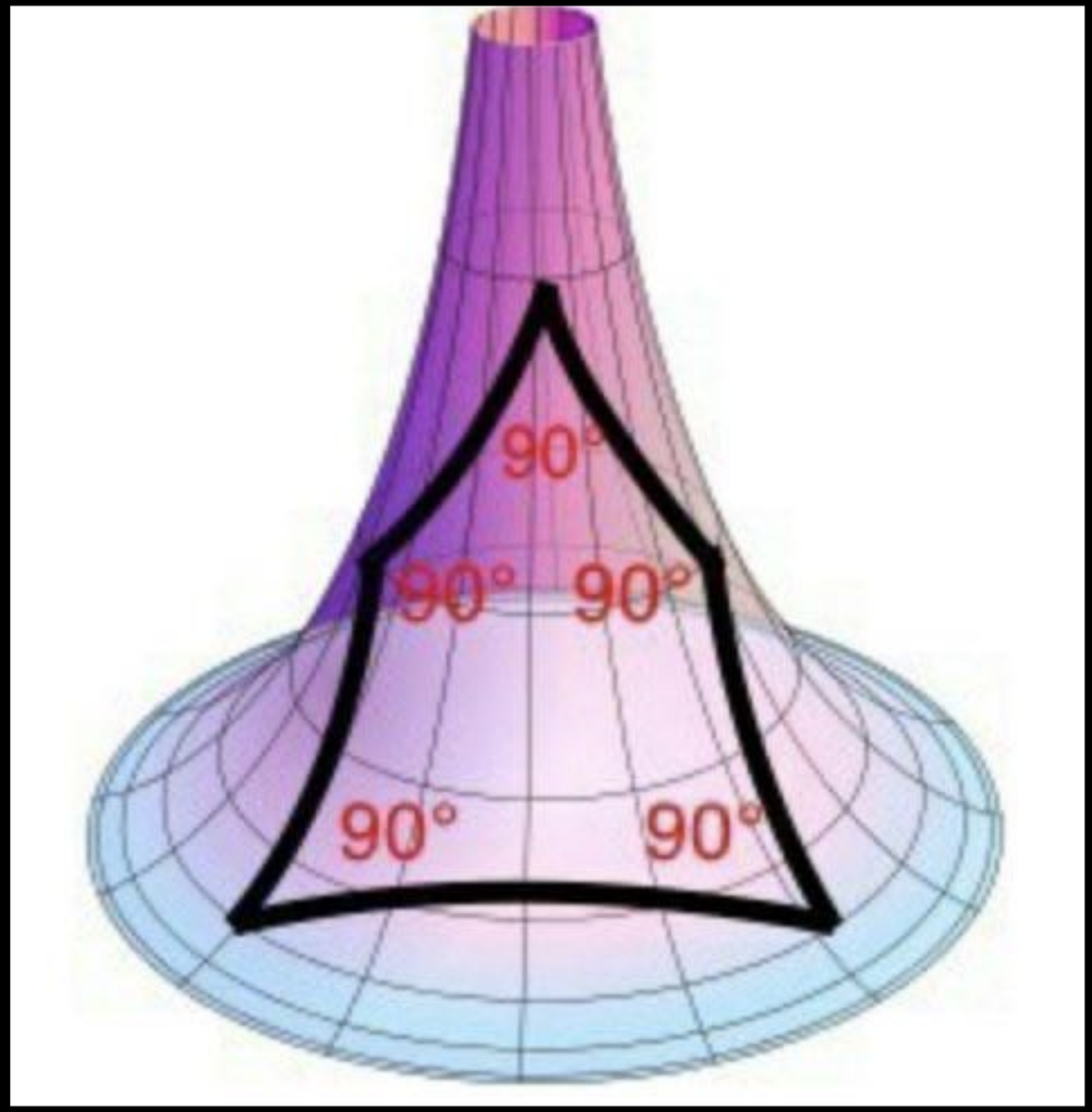
- 1823/1832, Bolyai’s (Hungarian) writes to his father “I have created a strange new universe”. His work “*The science absolute of space : independent of the truth or falsity of Euclid's axiom XI (which can never be decided a priori)*” is published in 1832.
- 1829 “*A Concise Outline of the Foundations of Geometry*” by Lobachevsky (Russian)
- 1848 Bolyai learns that Lobachevsky has published a similar piece. Their work is the basis for “hyperbolic geometry”
- 1905 Poincare describes his disk model of hyperbolic space and suggests that space might be hyperbolic.
- 1915 Einstein publishes “The field equations of gravitation”, describing space as non-euclidian.



What is Non-Euclidian geometry?

Among other things, this gives us five sides squares.

Actually, our space *is* hyperbolic due to the gravity of the Sun.



What is Non-Euclidian data?

- Vectorspaces like $\mathbb{R}^d$ have a euclidian metric (distance measure).
- But not all data follows these principles. For example:
  - There are hyperbolic vectorspaces where the parallel postulate does not hold.
  - Some data doesnt has a good notion of distance, or is irregular (eg with holes), so it isn’t even a vectorspace.

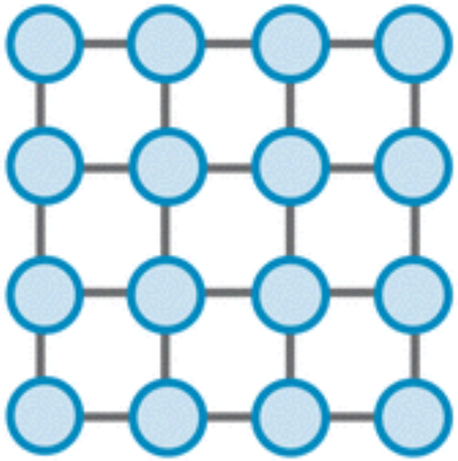
What is Non—Euclidian data?

Optionally, move beyond vectorspaces

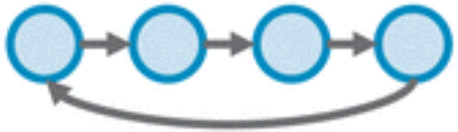
- Vectorspaces like $\mathbb{R}^d$ have a metric (distance measure).
- But not all data follows these principles. For example:
  - There are hyperbolic vectorspaces where the parallel postulate does not hold. There is still a metric (and a vectorspace).
  - Some data doesnt even has a good notion of distance, or is irregular (eg with holes). These are no longer vectorspaces, but topologies.

Regular Data Structures

Images

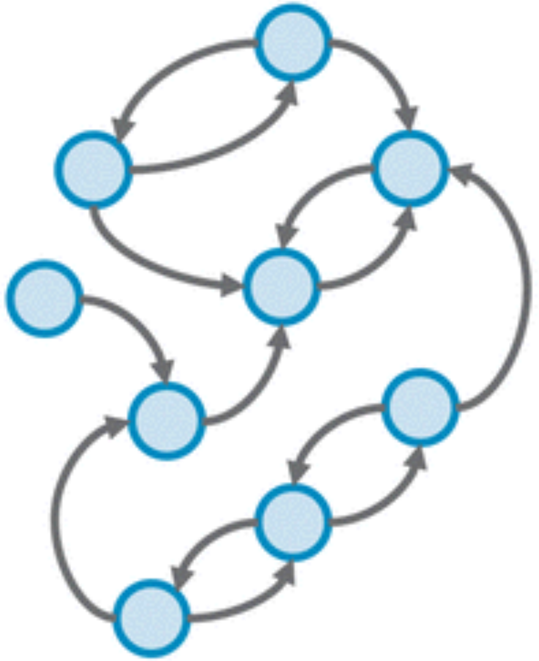


Time Series



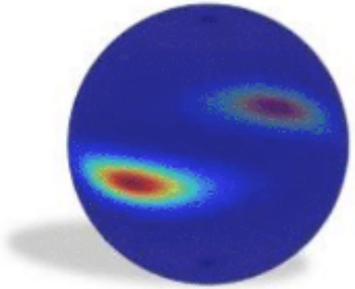
Irregular Data Structures

Social Networks
Sensor Feeds
Web Traffic
Supply Chains
Biological Systems
...



What is Non—Euclidian data?

Surfaces

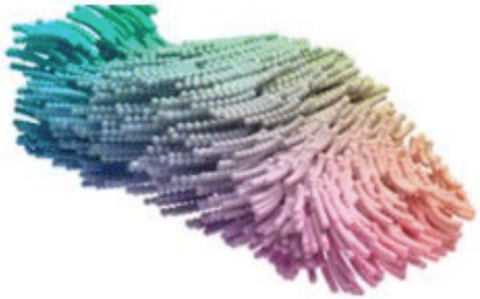


Distributions

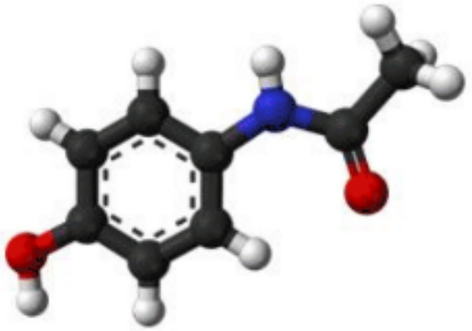
Graphs / Networks

Functions on Manifolds

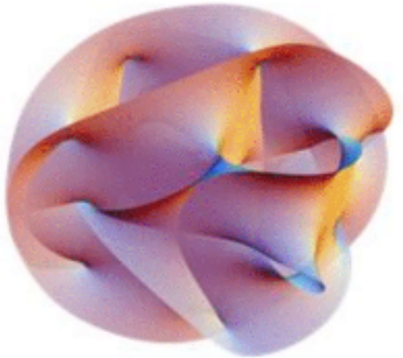
Hyperbolic spaces



Hyper-surfaces



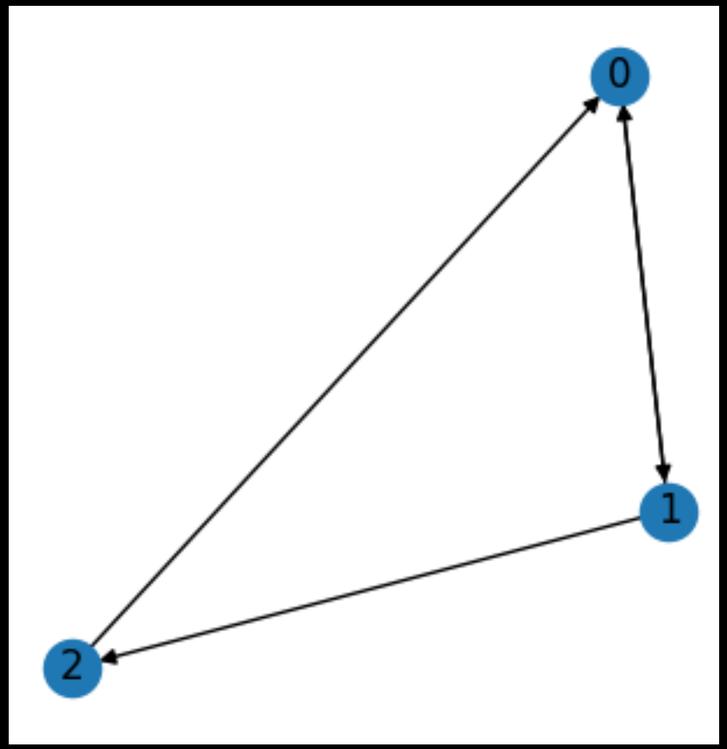
Molecules



General manifolds

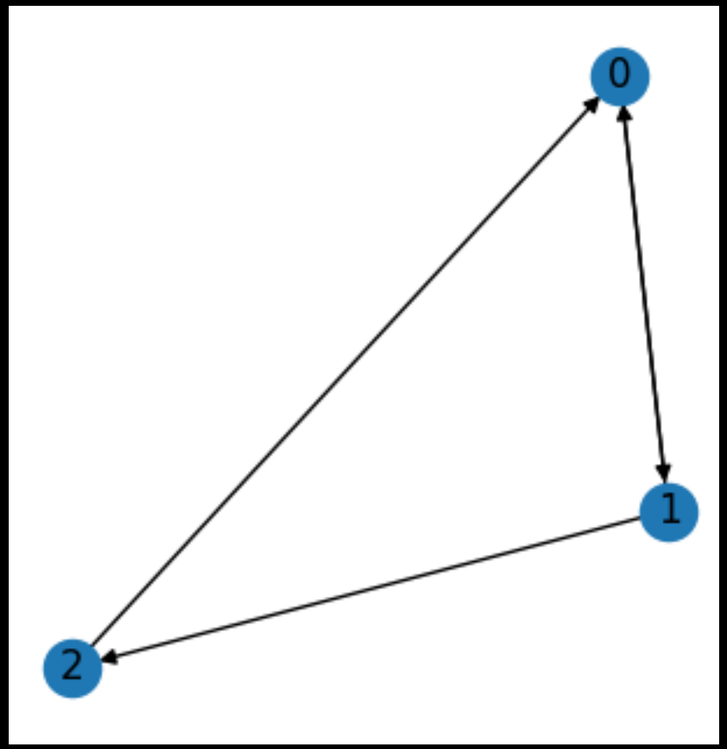
What is a Graph?

- A Graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ is defined by:
  - A set of nodes $\mathcal{V} = \{v_1, \ldots, v_n\}$
  - A set of edges between nodes $\mathcal{E} = \{(v_i, v_j) \,|\, v_i, v_j \in \mathcal{V}\}$



What is a Graph?

- The adjecency matrix A has a 1 on every position where there is and edge:
  $A[i,j] = 1$ if $e_{i,j} \in \mathscr{E}$
- Excercise: let's draw A for this graph!



Some statistics on graphs

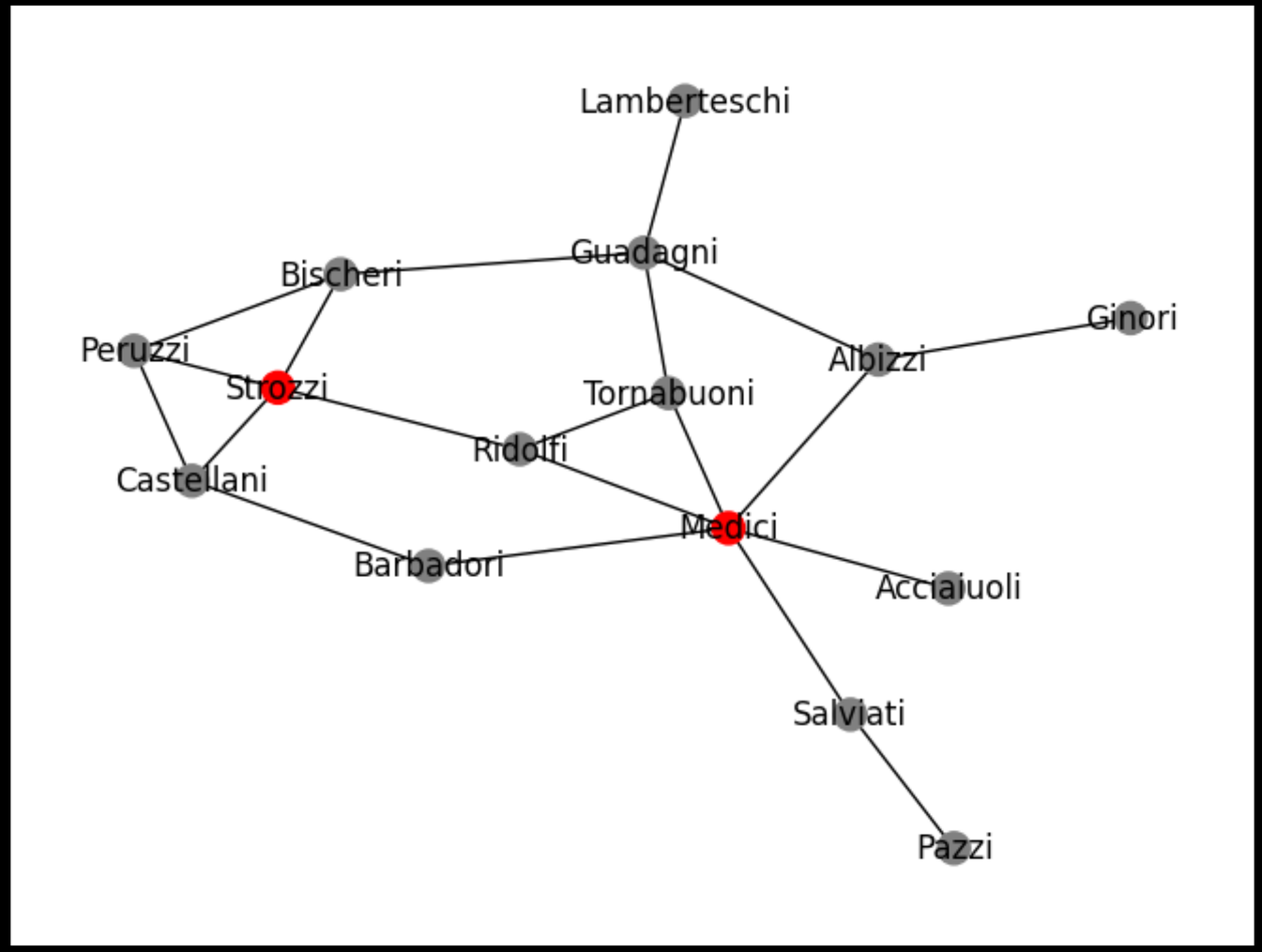
- Degree: the number of edges for a node $d_u = \Sigma_{v \in V} A[u, v]$
- The degree matrix D has on each diagonal element the degree of the note: $D[i, i] = d_i$
- Betweenness centrality is the sum of the fraction of all shortest paths through v: $cb(v) = \Sigma_{s,t \in V} \frac{\sigma(s, t \mid v)}{\sigma(s, t)}$ with $\sigma(s, t)$ the number of shortest (s,t) paths and $\sigma(s, t \mid v)$ the paths through v.

Florentine Families

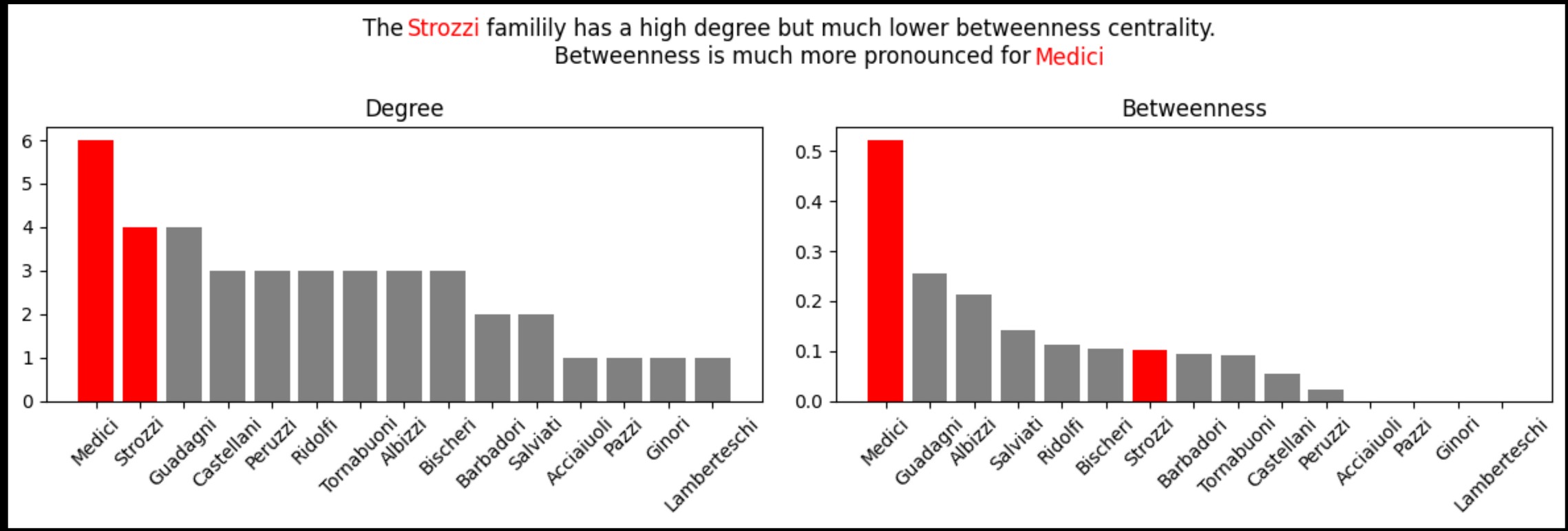
Renaissance Florentine families around 1430, collected by John Padgett from historical documents.

The graph shows marriage alliances.

The data include families who were locked in a struggle for political control. Two factions were dominant in this struggle: one revolved around the Medicis, the other around the Strozzis.



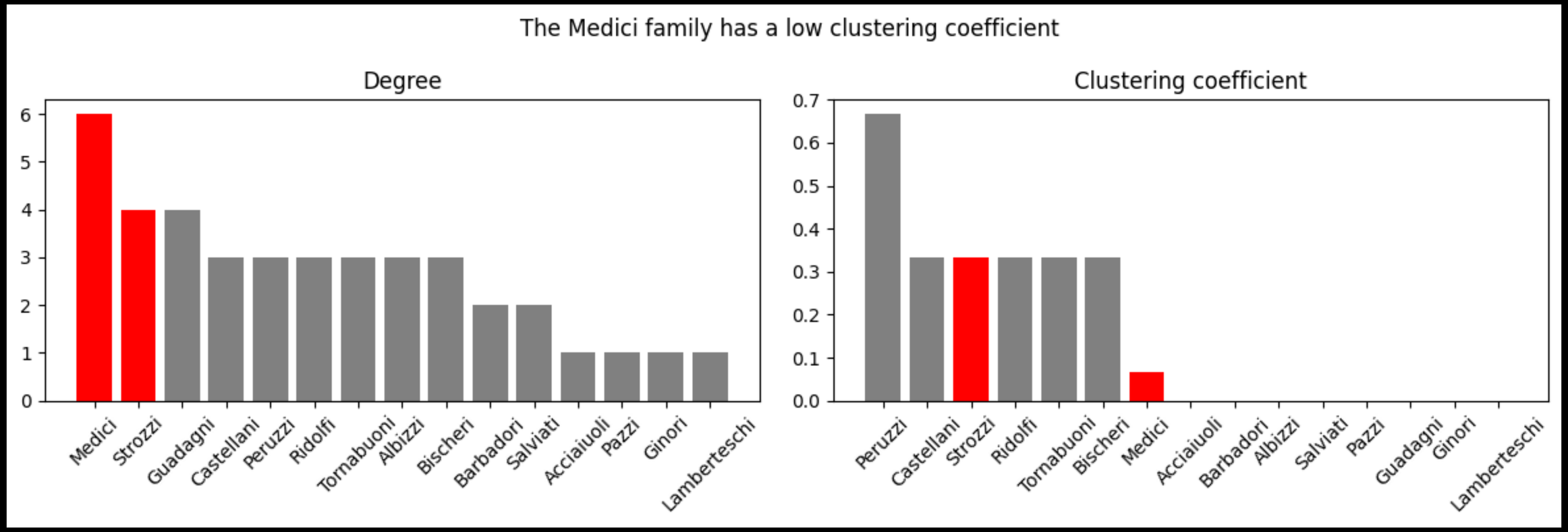
Florentine Families



Some statistics on graphs

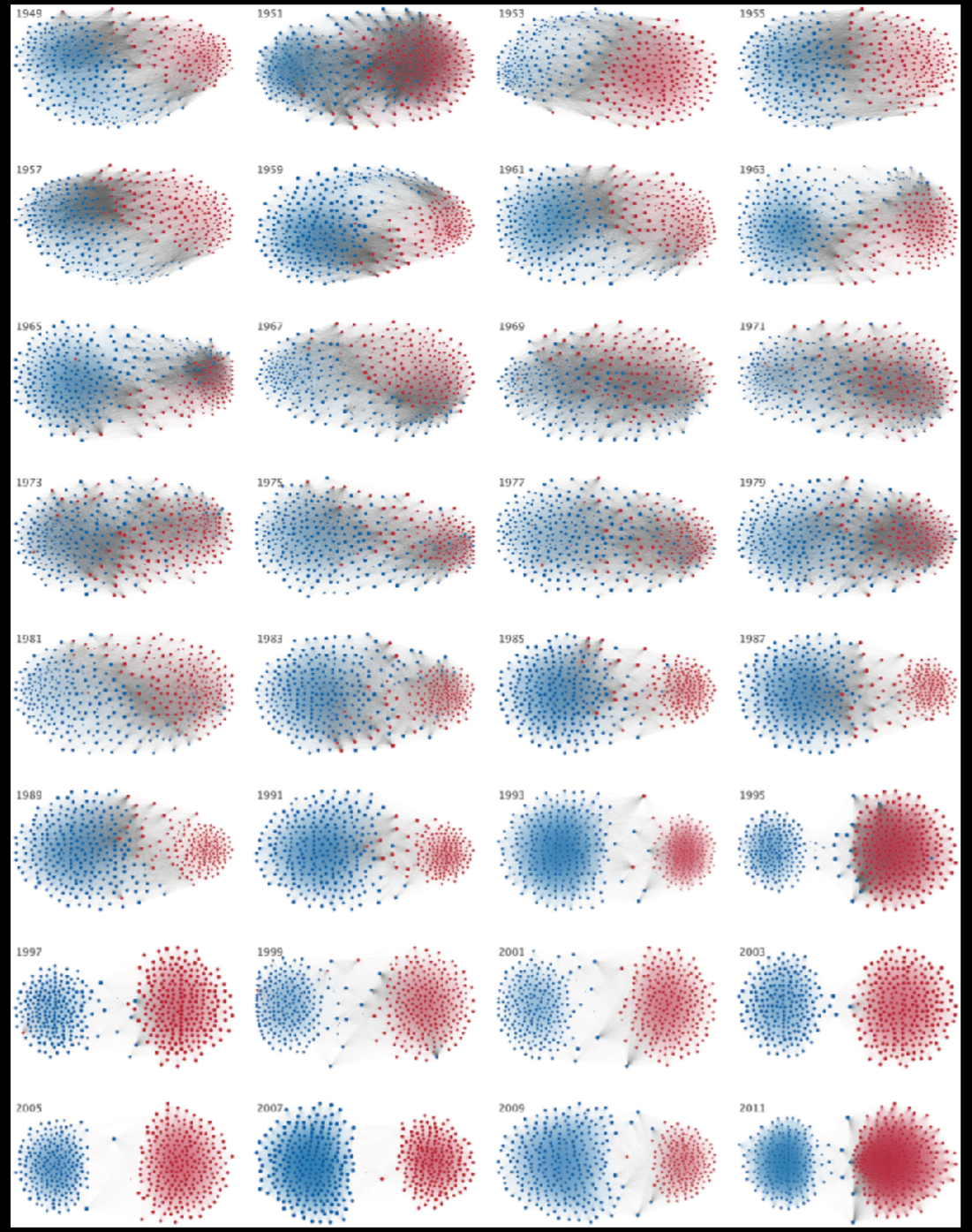
- Triangles: if your friend are also friends
- Clustering coefficient: the fraction of possible triangles.
- This can be a very relevant metric: e.g. there is a correlation between (lack of) triades and depression!

Florentine Families



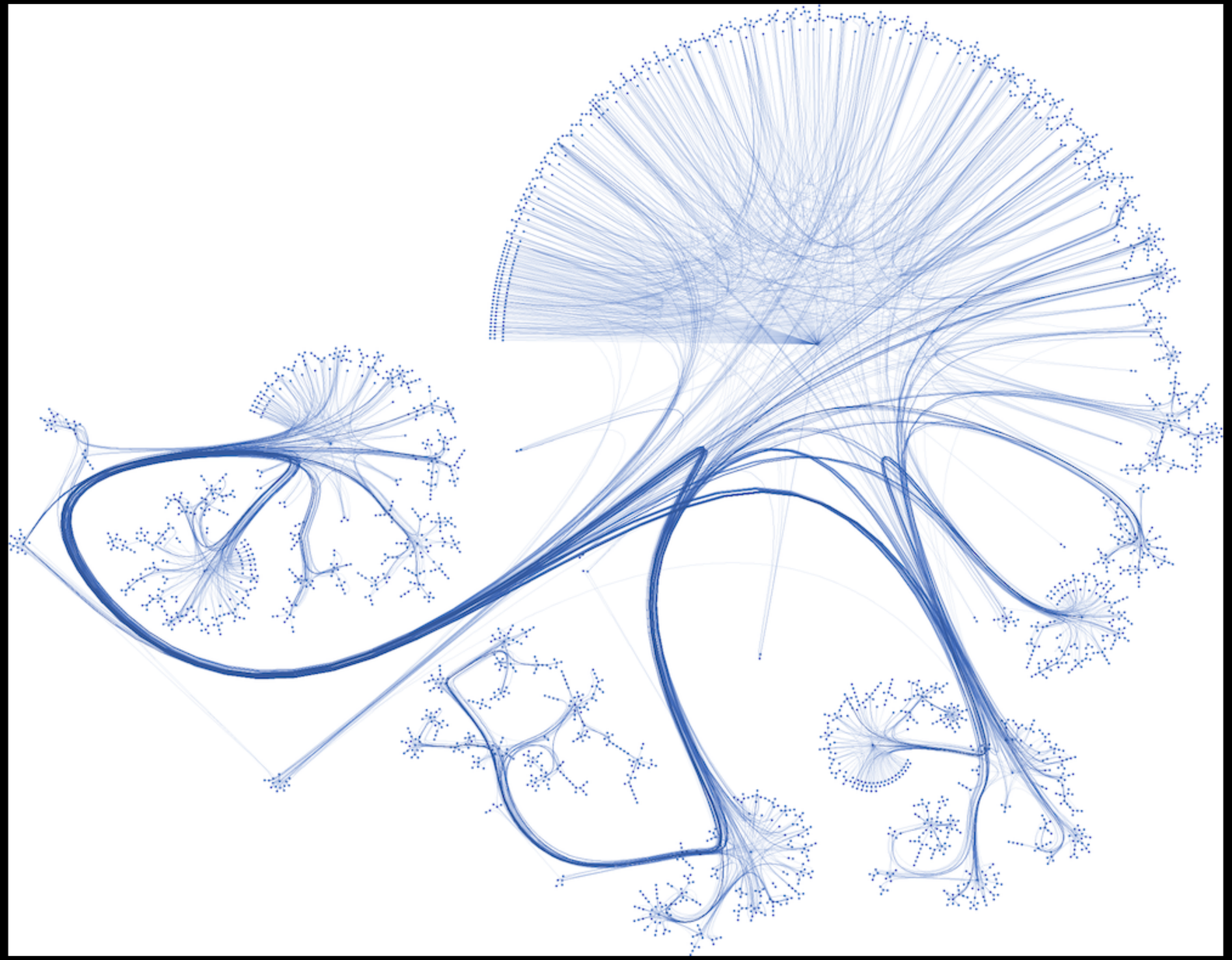
See US Congress polarize over the past 60 years

See how likely the House of Representatives' Democrats (in blue) and Republicans (in red) are to vote with their own party, or to cross party lines.

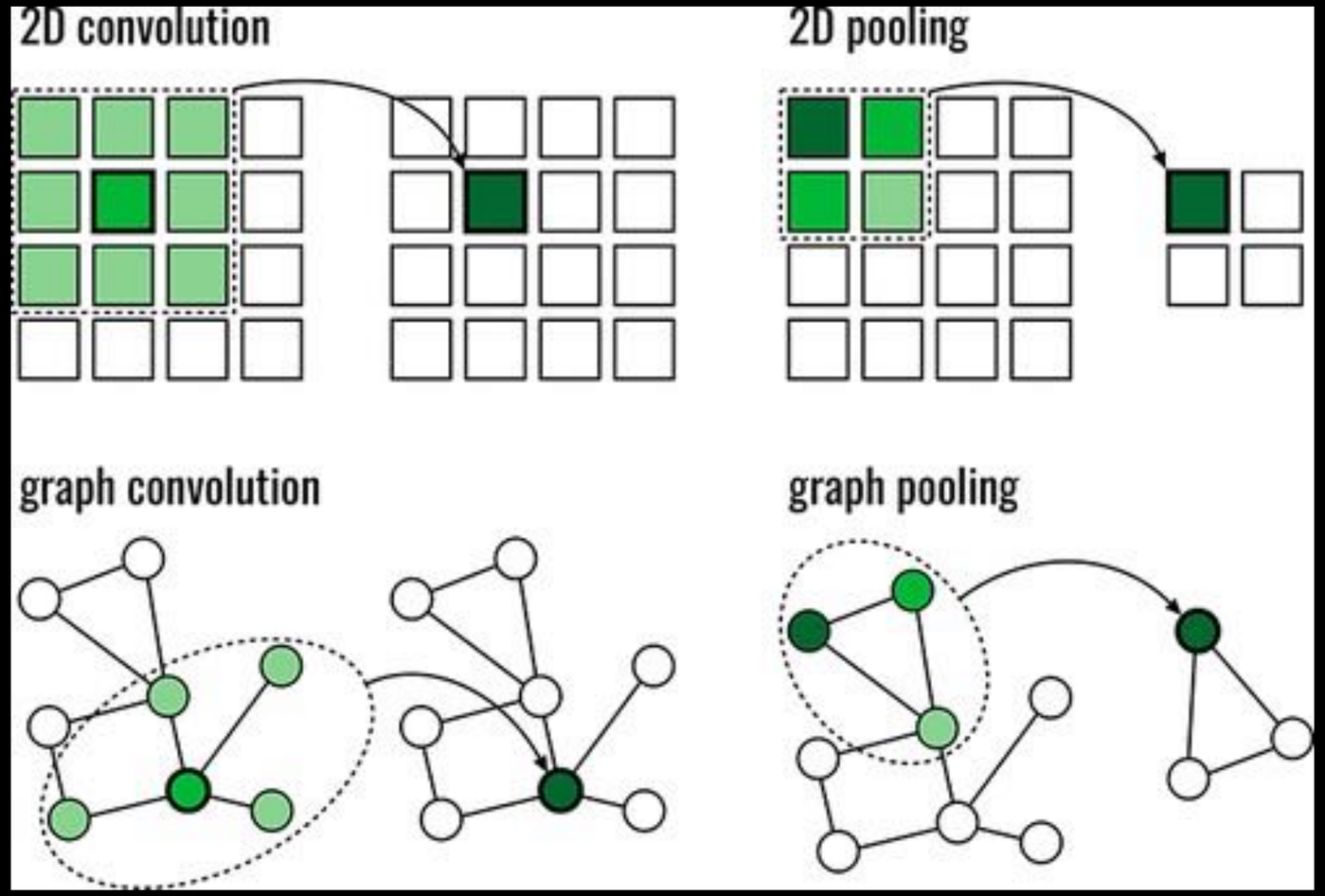


Cora dataset

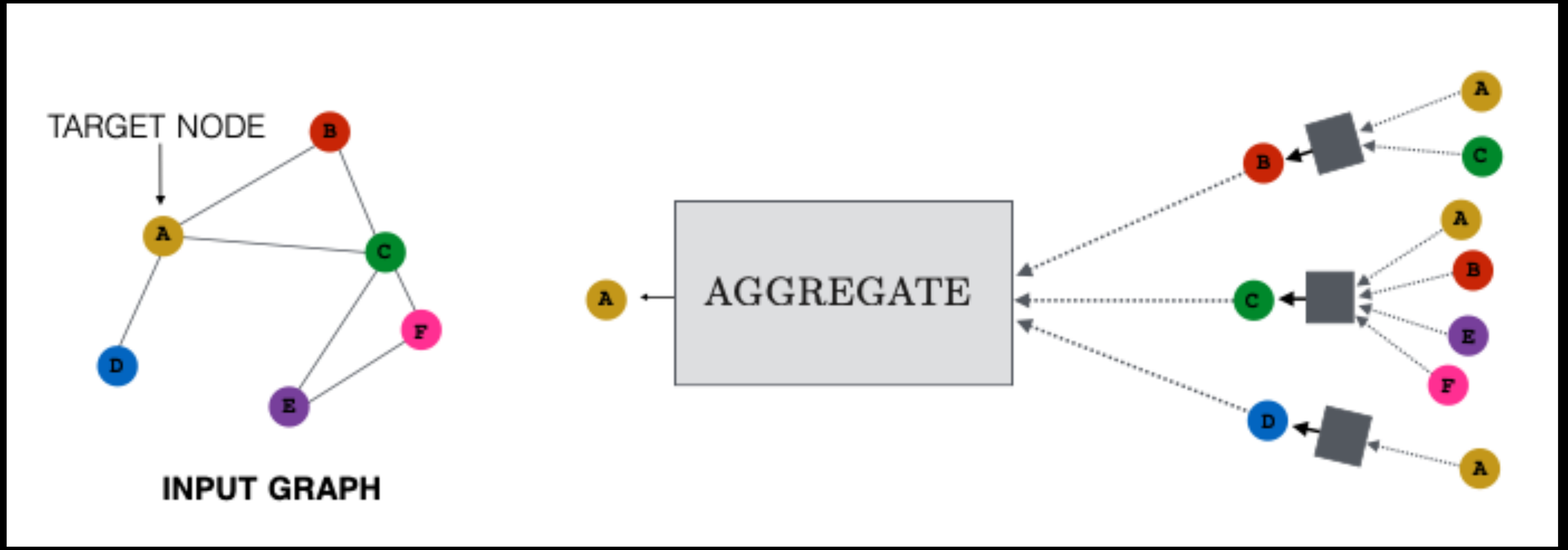
- 2708 scientific publications
- classified into one of seven classes.
- The citation network consists of 5429 links.
- Each publication in the dataset is described by a 0/1-valued word vector indicating the absence/presence of the corresponding word from the dictionary. The dictionary consists of 1433 unique words



Graph convolutions



Graph convolutions



Graph convolutions

- At each iteration, every node aggregates information from its local neighborhood
- After k convolutions, each node contains information from its k-hop neighborhood

Graph convolutions

Basic message passing

- $h_u^{(k)}$ is the embedding of node u after k convolutions.
- $W_{self}, W_{neigh} \in \mathbb{R}^{d_k \times d_{k-1}}$ are trainable weights

$$\mathbf{h}_u^{(k)} = \sigma\left(\mathbf{W}_{\text{self}}^{(k)}\mathbf{h}_u^{(k-1)} + \mathbf{W}_{\text{neigh}}^{(k)} \sum_{v \in \mathcal{N}(u)} \mathbf{h}_v^{(k-1)} + \mathbf{b}^{(k)}\right)$$