A few announcements…

CSC630: Advanced topics in interactive data analysis

- catalog advertises as “Advanced topics in Software Systems”
- Research seminar
  - all paper readings, discussions, presentations, projects

CSC630: Advanced topics in interactive data analysis

We need **interactive interfaces**: computer programs to help us build models and visualizations progressively.

We will spend most of the class time discussing recent research results in the area, in the form of conference and journal papers.

How can we build these systems? How do we know they work? Can they mislead us, and how? How do they fit with the rest of the data analysis infrastructure?

We will cover a mix of visualization, algorithms, systems, data mining, machine learning, and whatever other computer science topics become necessary as our discussion progresses.

Students will be assessed on their class participation, paper summary reports, and research projects.

Calendar adjustments

Apr 21st: Trees, Graphs, Hierarchies
Apr 23rd: Spatial Data: heatmaps, contour plots, vector fields
Apr 28th: Spatial Data: heatmaps, contour plots, vector fields
Apr 30th: Spatial Data: heatmaps, contour plots, vector fields
May 5th: Methods for large data; binning, sampling
May 7th: Uncertainty/Probabilistic Data
May 12th: Catchup?
May 14th: Final Presentation

Calendar adjustments

Apr 21st: Trees, Graphs, Hierarchies
Apr 23rd: Spatial Data: heatmaps, contour plots, vector fields
Apr 28th: Spatial Data: heatmaps, contour plots, vector fields
Apr 30th: Methods for large data; binning, sampling
May 5th: Uncertainty/Probabilistic Data
May 7th: Dead Day
May 12th: Finals week
May 14th: Finals week

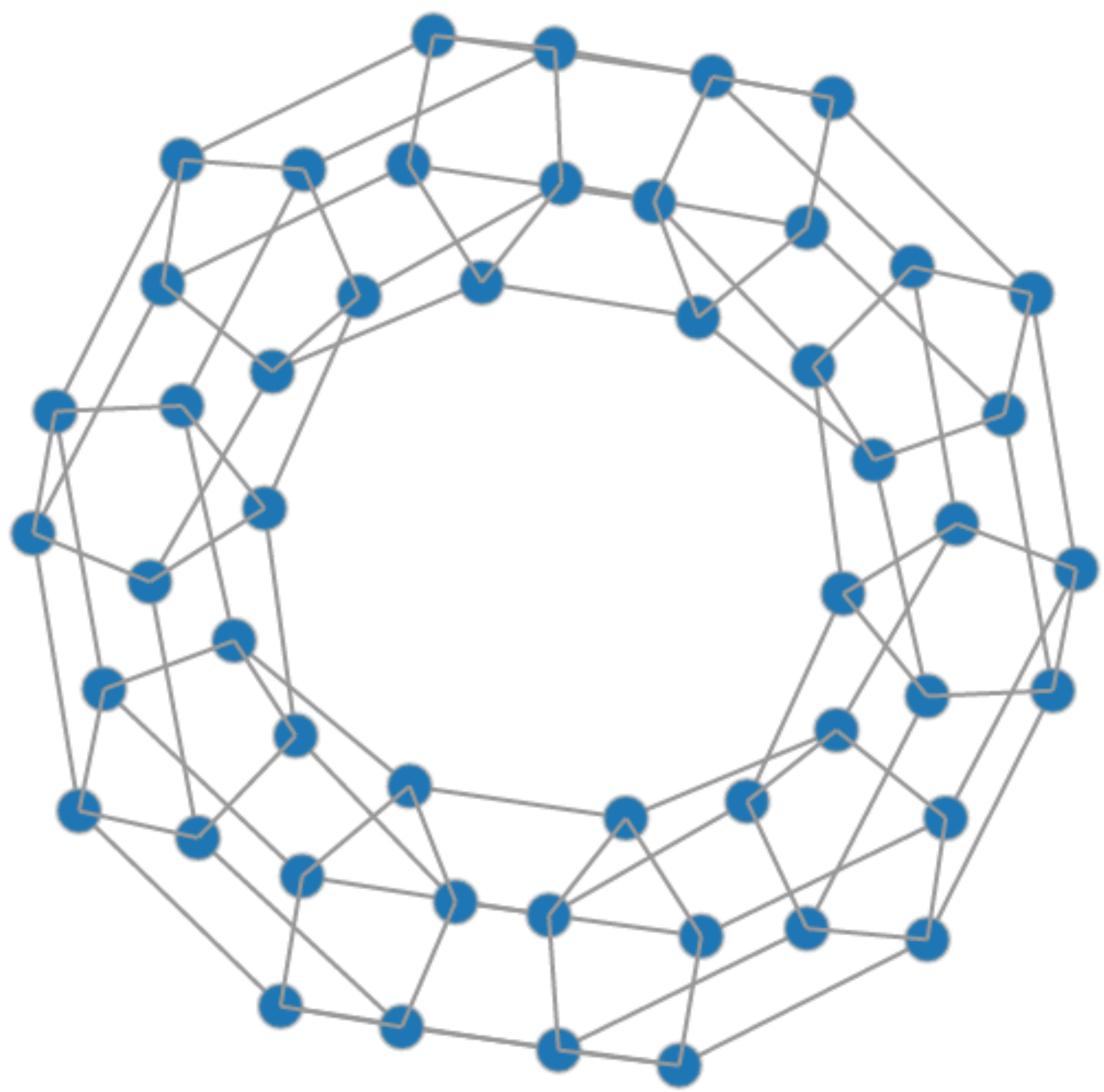
Calendar adjustments

- Necessary change: you will present your final project to me outside of regular class schedules, in order to have enough time to work on it.
- You can present at any time from here to May 14th.

Graphs

CS444/544

Node-link diagrams



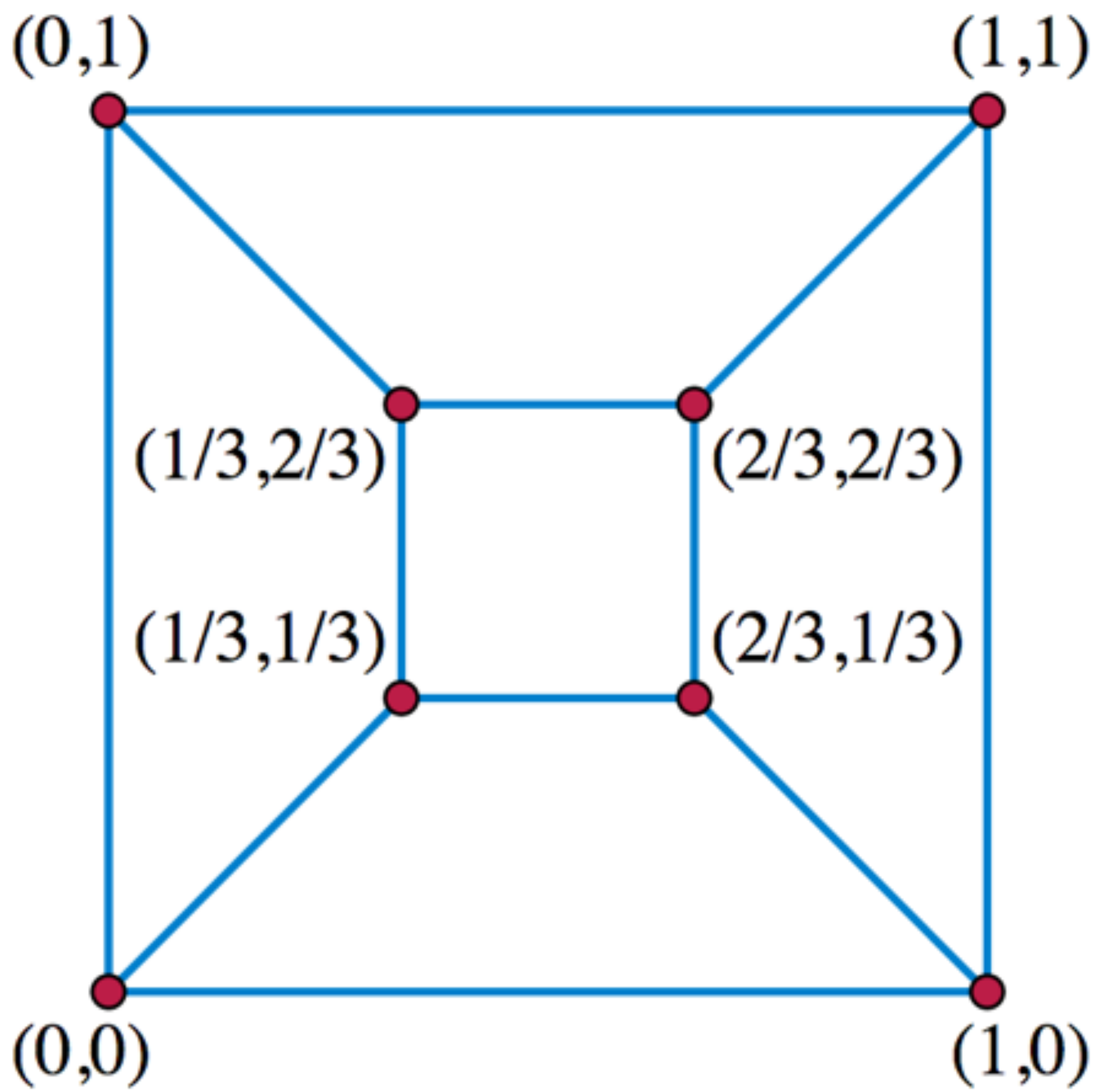
http://christophermanning.org/gists/1703449/#/%5B10%5D50/1/0

Starting simple: planar 3-vertex connected graphs (what?)

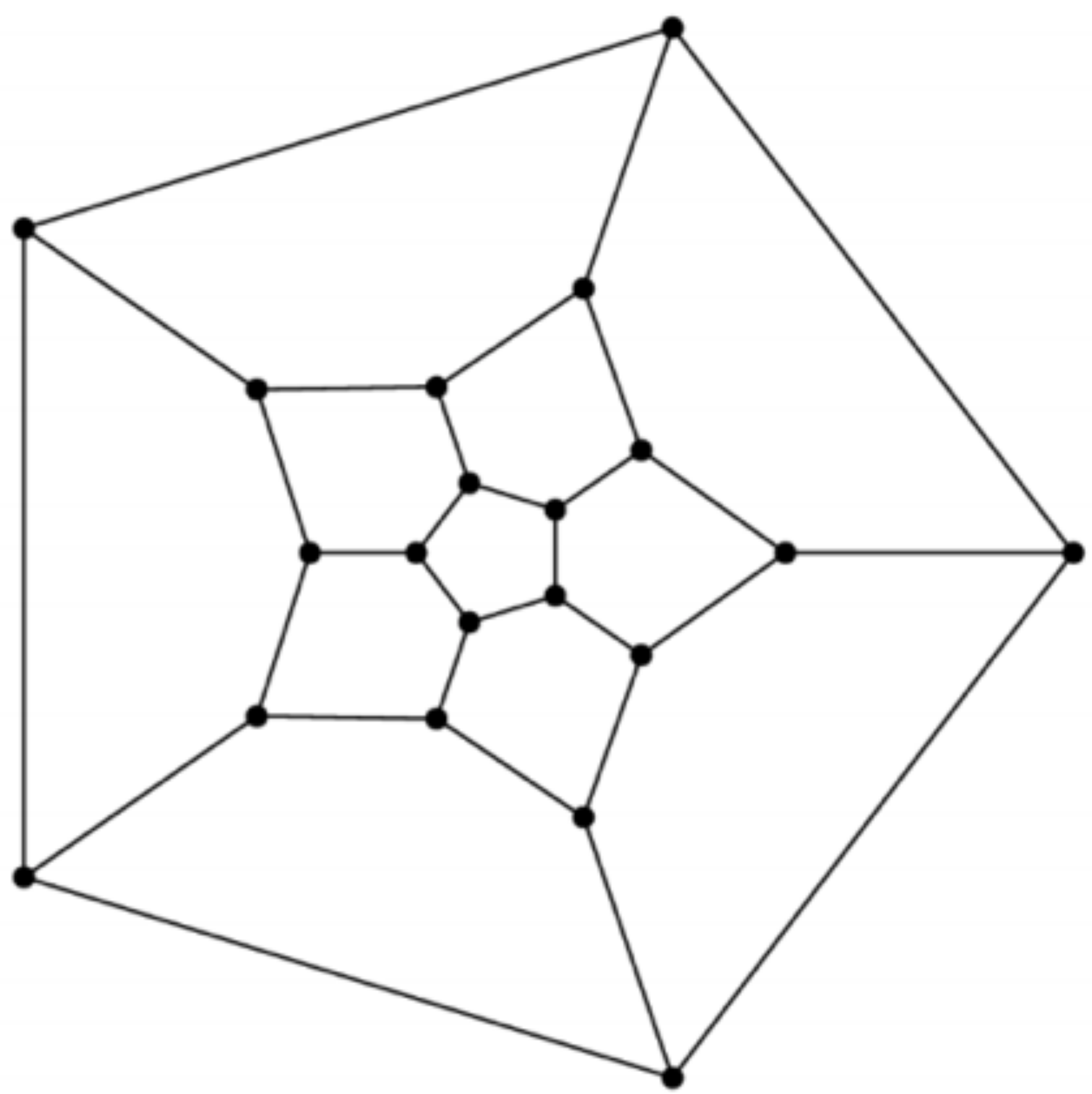
Tutte Embedding

- Each node should be the average of its neighbors
  - Aside from the boundary, which is user-specified
- This gives a linear system
- Theorem: **if graph is planar, embedding is crossing-free**

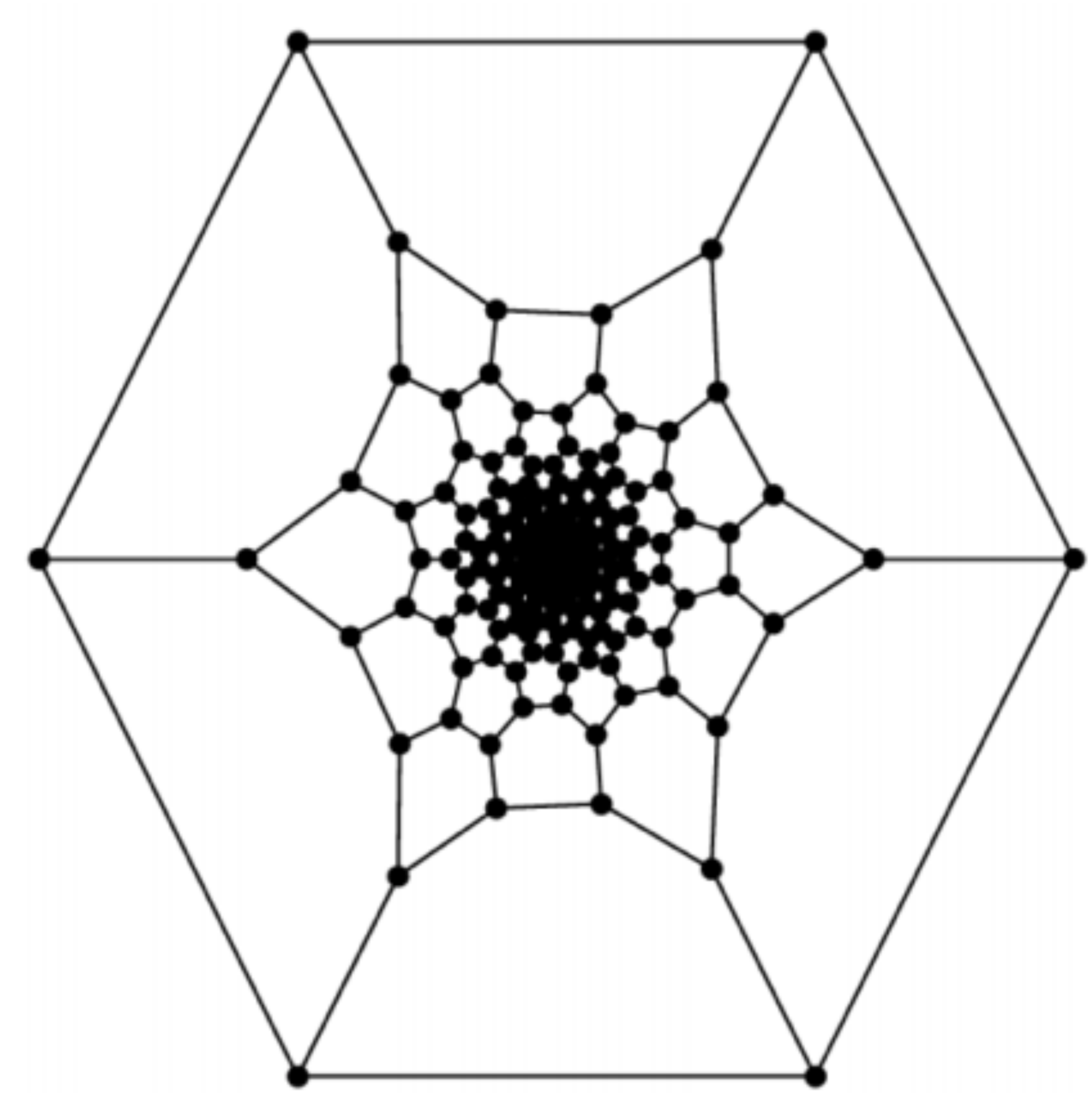
Tutte Embedding



Downsides



Dodecahedron



$Le(C60)$

http://www.cs.arizona.edu/~kpavlou/Tutte_Embedding.pdf

Force-directed Layouts

- Intuition: define “forces” on “physical objects”, initialize positions randomly, let the system settle

http://bl.ocks.org/mbostock/4062045

- Need to define what forces are, and what physical objects are

Force-directed Layouts

- We want edges to be neither too small or too large
  - Physical analogy: **Springs compress or expand to achieve ideal length**
- We don't want vertices to bunch up together
  - Physical analogy: **Electric charges with the same sign don't bunch up**

Force-directed Layouts

- Force per edge: $f_E(d) = C_E \times (d - L)$

Force-directed Layouts

- Force per vertex pair: $f_V(d) = C_V \times \frac{m_1 m_2}{d^2}$

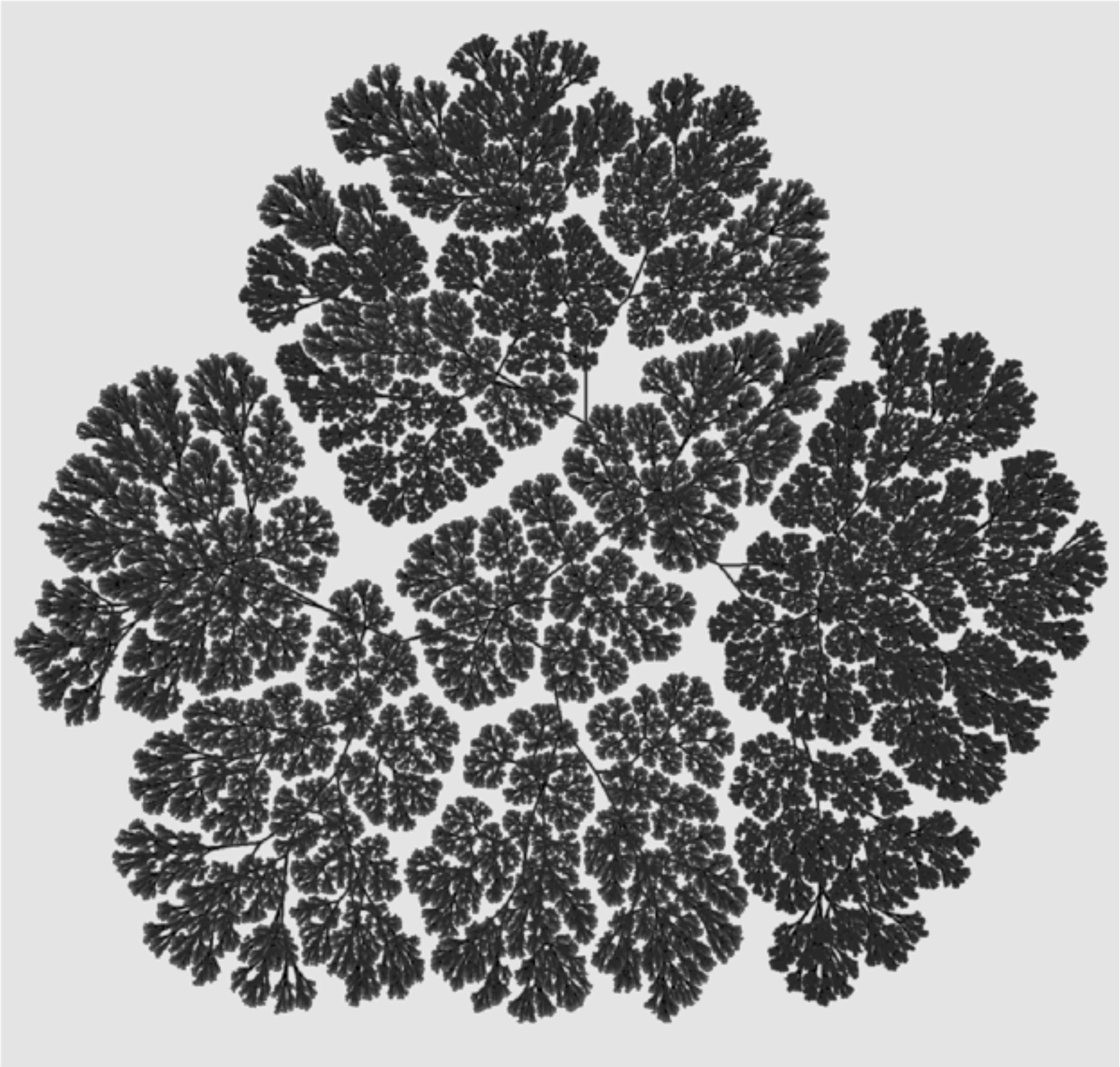
Force-directed Layouts

- Algorithm:
- For each vertex, determine all forces that apply to it,
    - Edges $f_E(d) = C_E \times (d - L)$
    - vertices $f_V(d) = C_V \times \frac{m_1 m_2}{d^2}$
- compute direction of movement, move small amount in those directions
    - iterate until convergence

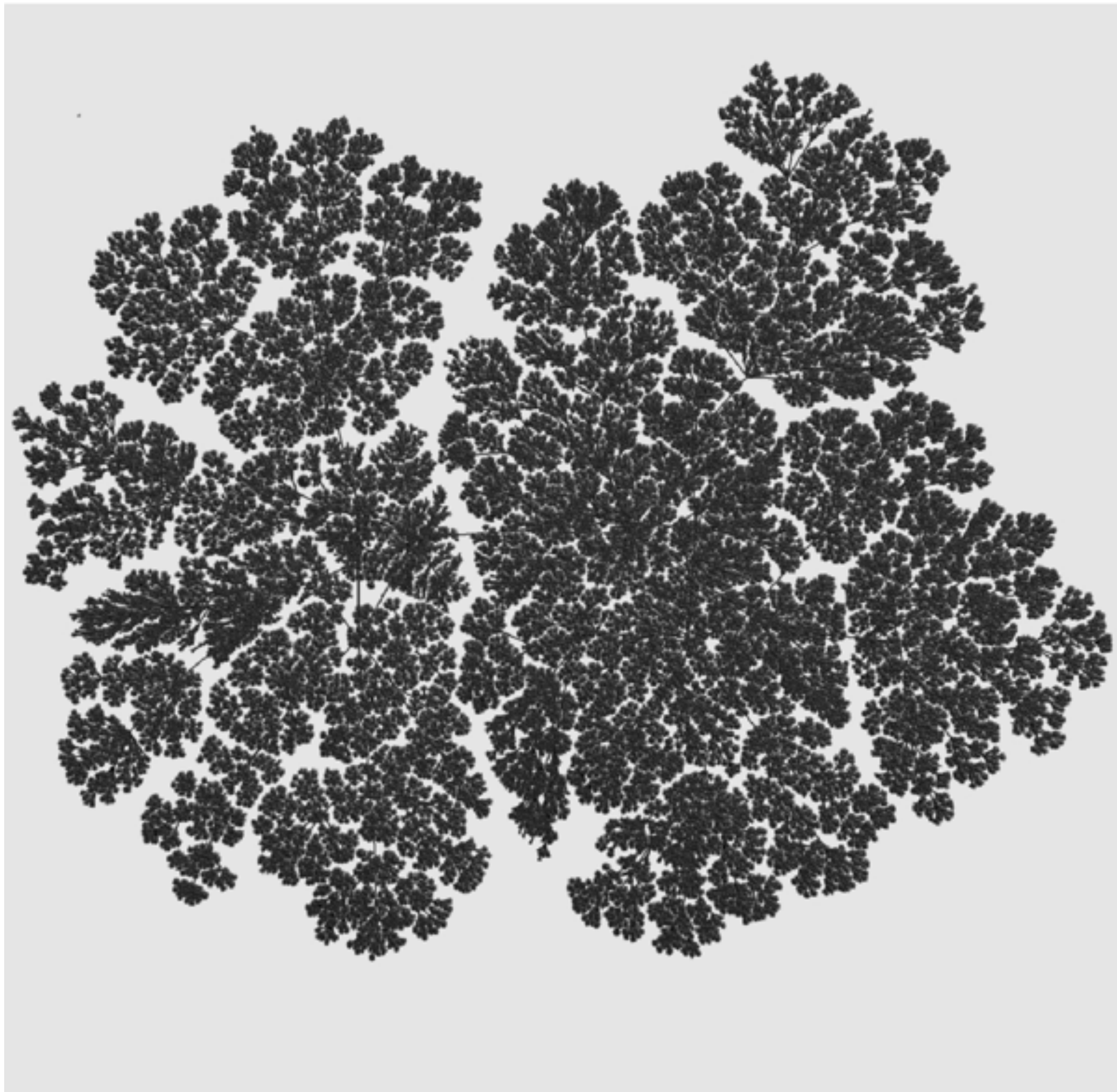
Downsides

- Requires $O(|V|^2)$ work per step
  - Faster algorithms exist: Barnes-Hut, multipole methods, etc.
- For large graphs, result is not very informative

Downsides



Downsides

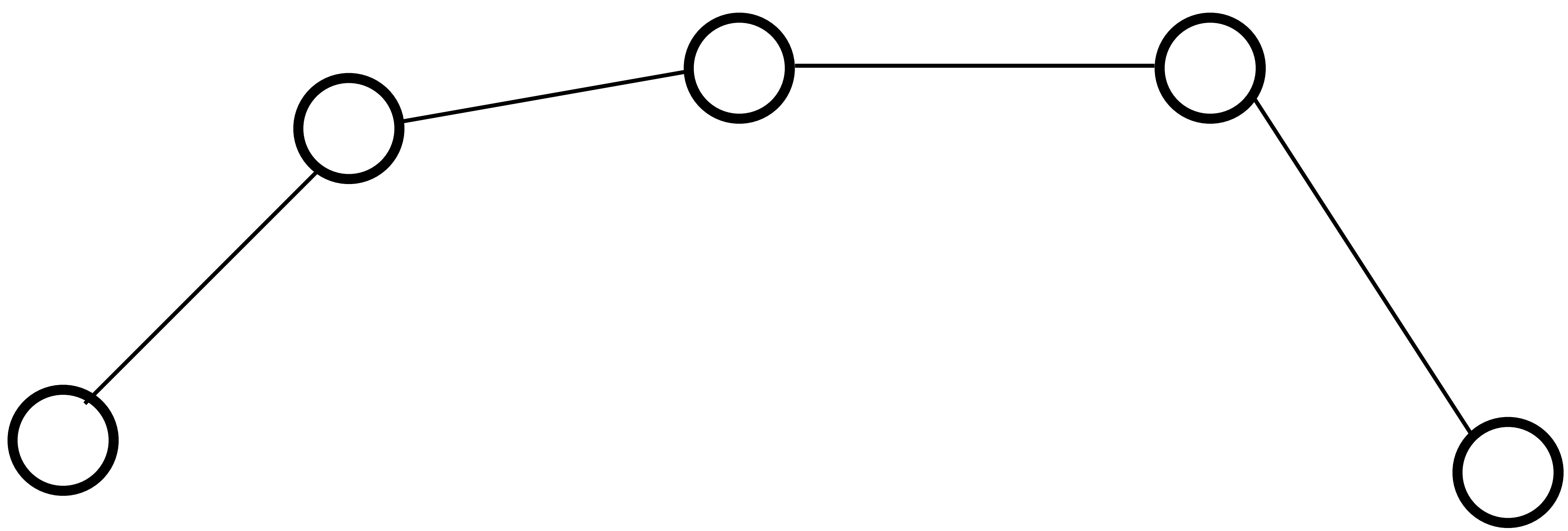


Metric Embeddings

- Use **global properties** of the graph instead of only local interactions
- Specifically, graph **distances**

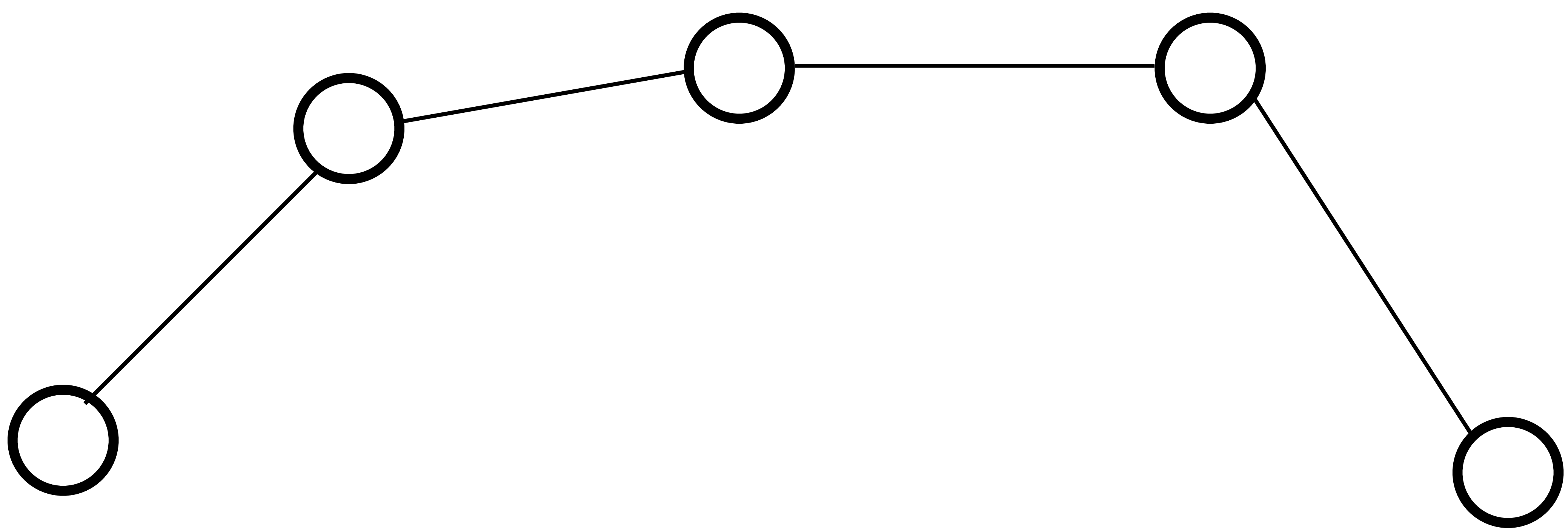
Metric Embeddings

- Graph distances can be used to define “forces”
  - Encode directly that **far away vertex pairs should be placed far from one another**



Metric Embeddings

$$E(X) = \sum_{i,j} (d(i,j) - |X_i - X_j|)^2$$

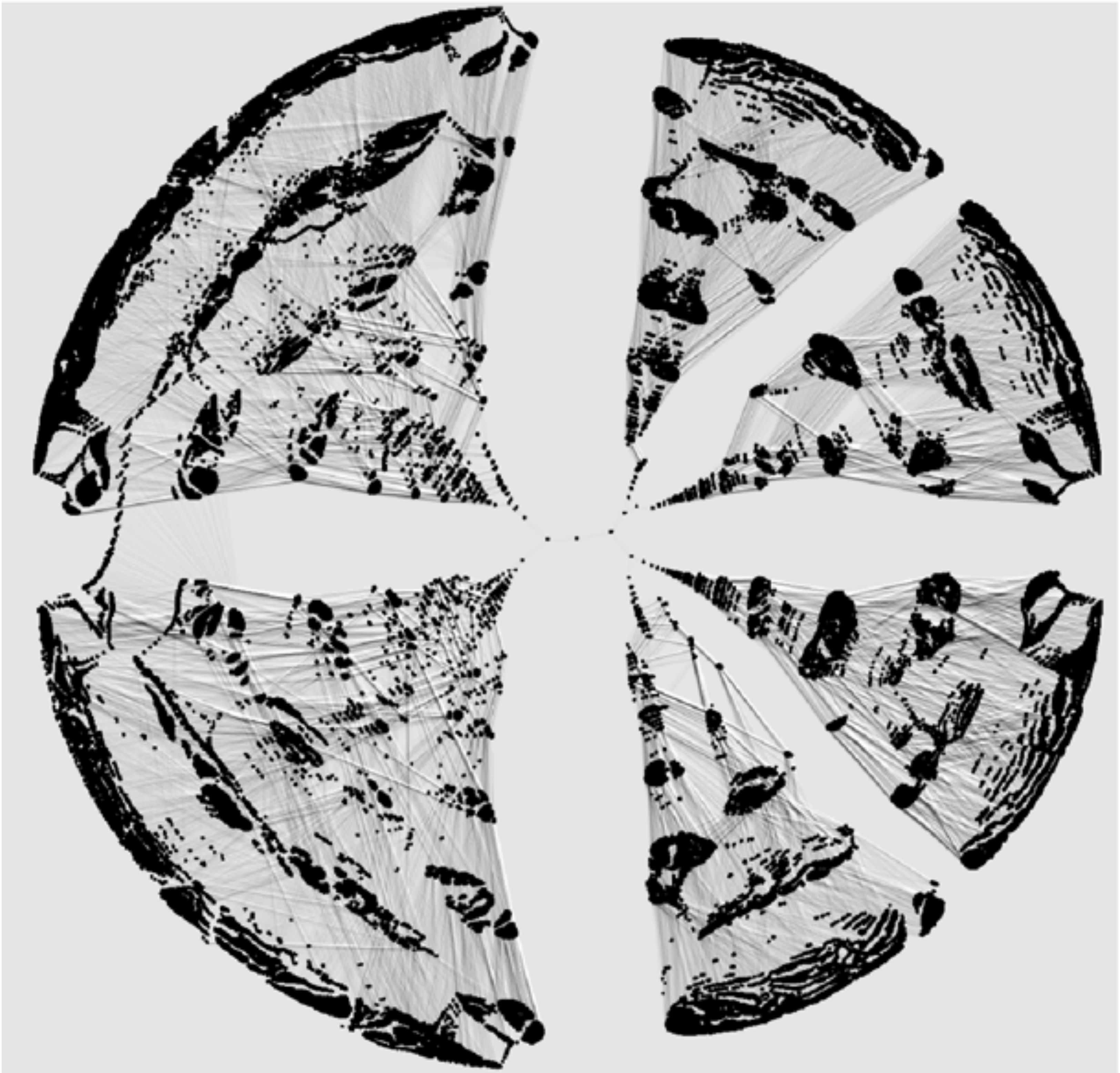


Metric Embeddings

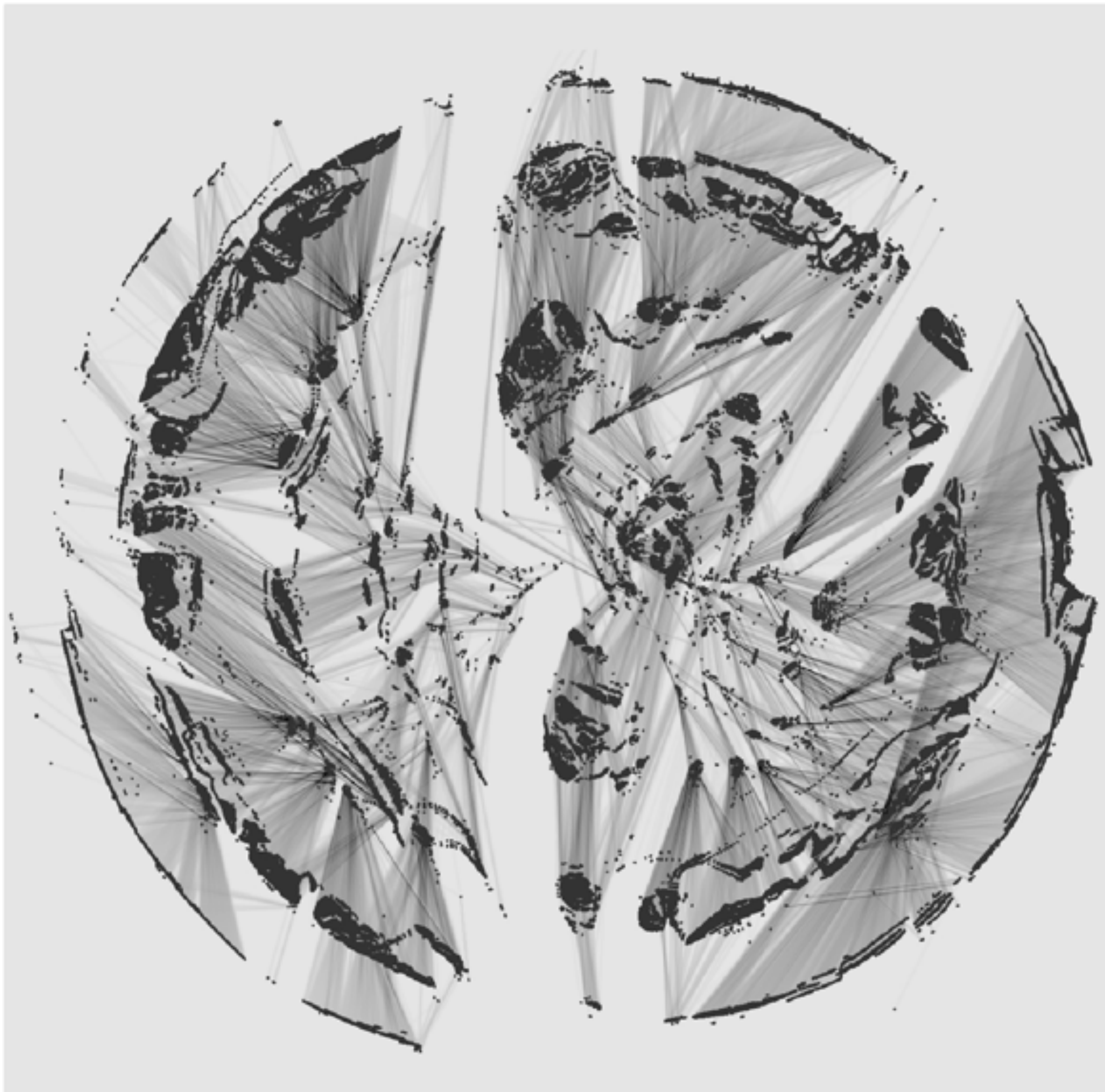
$$E(X) = \sum_{i,j} (d(i,j) - |X_i - X_j|)^2$$

- Our old friend, dimensionality reduction!

Metric Embeddings

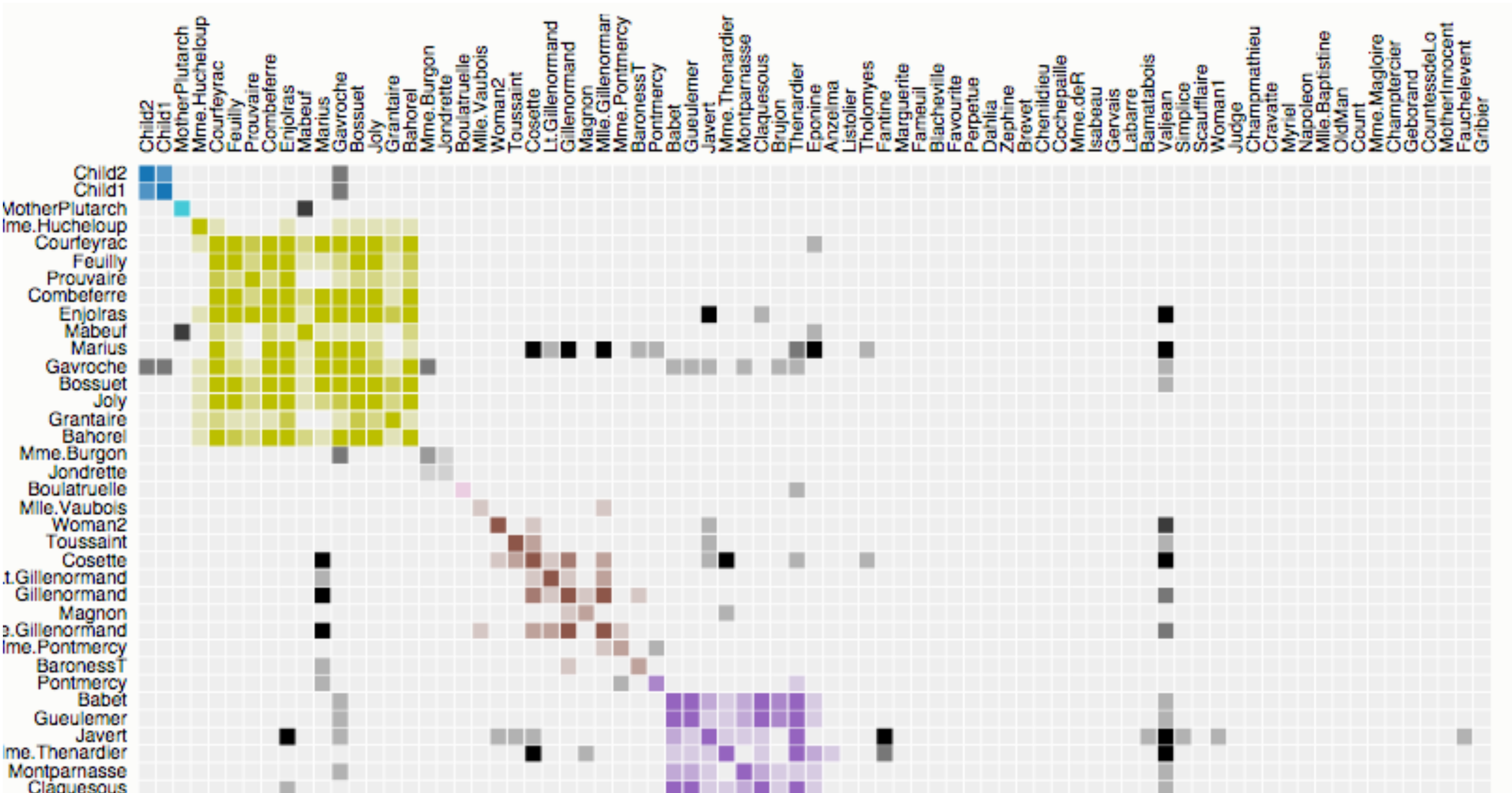


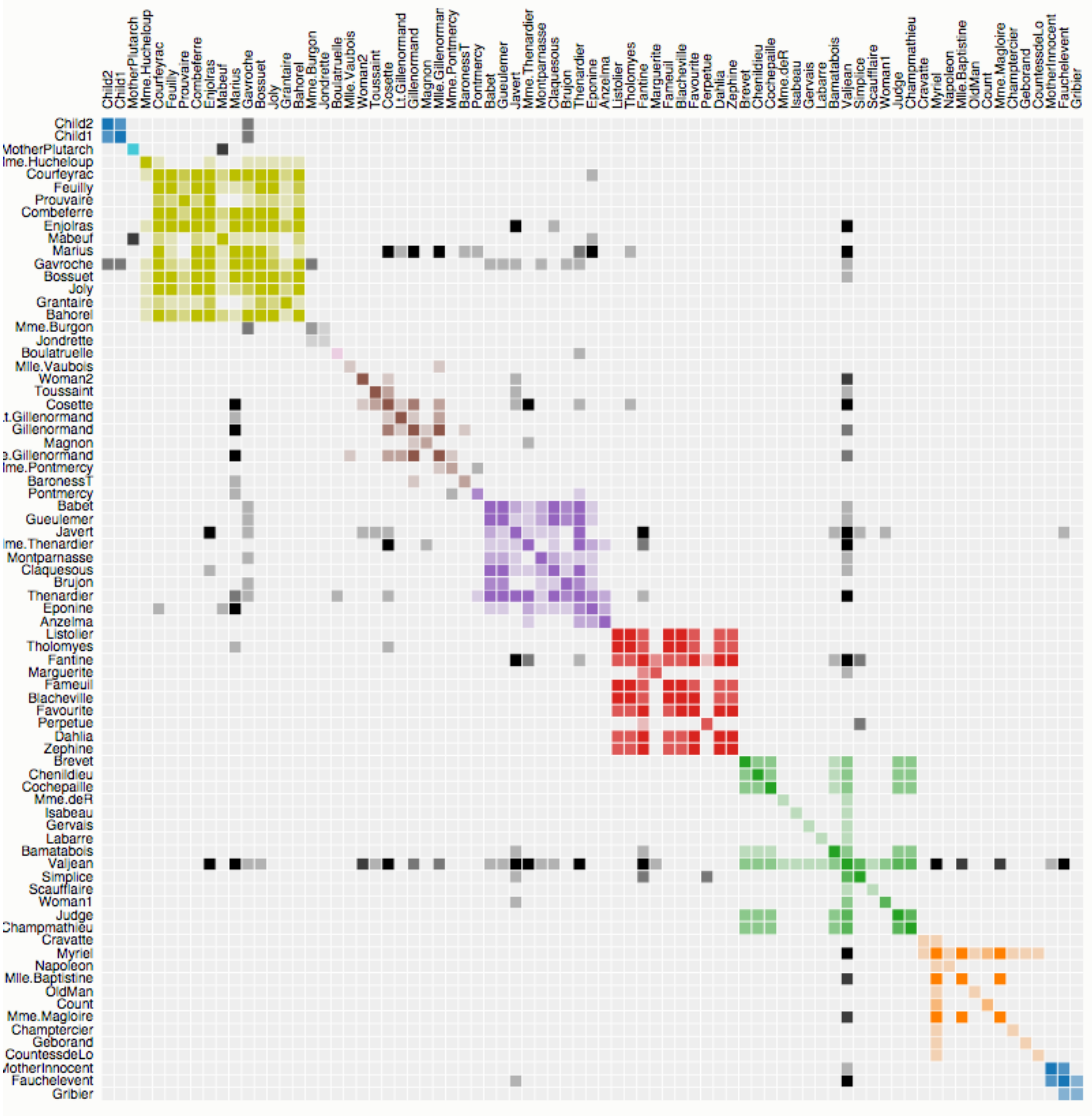
Metric Embeddings



Matrix Diagrams

http://bost.ocks.org/mike/miserables/




Child2
Child1
MotherPlutarch
Mme.Hucheloup
Courfeyrac
Feuilly
Prouvaire
Combeferre
Enjolras
Mabeuf
Marius
Gavroche
Bossuet
Joly
Grantaire
Bahorel
Mme.Burgon
Jondrette
Boulatruelle
Mlle.Vaubois
Woman2
Toussaint
Cosette
Lt.Gillenormand
Gillenormand
Magnon
Mlle.Gillenormand
Mme.Pontmercy
BaronessT
Pontmercy
Babet
Gueulemer
Javert
Mme.Thenardier
Montparnasse
Claquesous
Brujon
Thenardier
Eponine
Anzelma
Listolier
Tholomyes
Fantine
Marguerite
Fameuil
Blacheville
Favourite
Perpetue
Dahlia
Zephine
Brevet
Chenildieu
Cochepaille
Mme.deR
Isabeau
Gervais
Labarre
Bamatabois
Valjean
Simplice
Scaufflaire
Woman1
Judge
Champmathieu
Cravatte
Myriel
Napoleon
Mlle.Baptistine
OldMan
Count
Mme.Magloire
Champtercier
Geborand
CountessdeLo
MotherInnocent
Fauchelevent
Gribier

Upsides

- Easy to define for directed and undirected graphs
- Easy to compute
- Easy to incorporate edge attributes