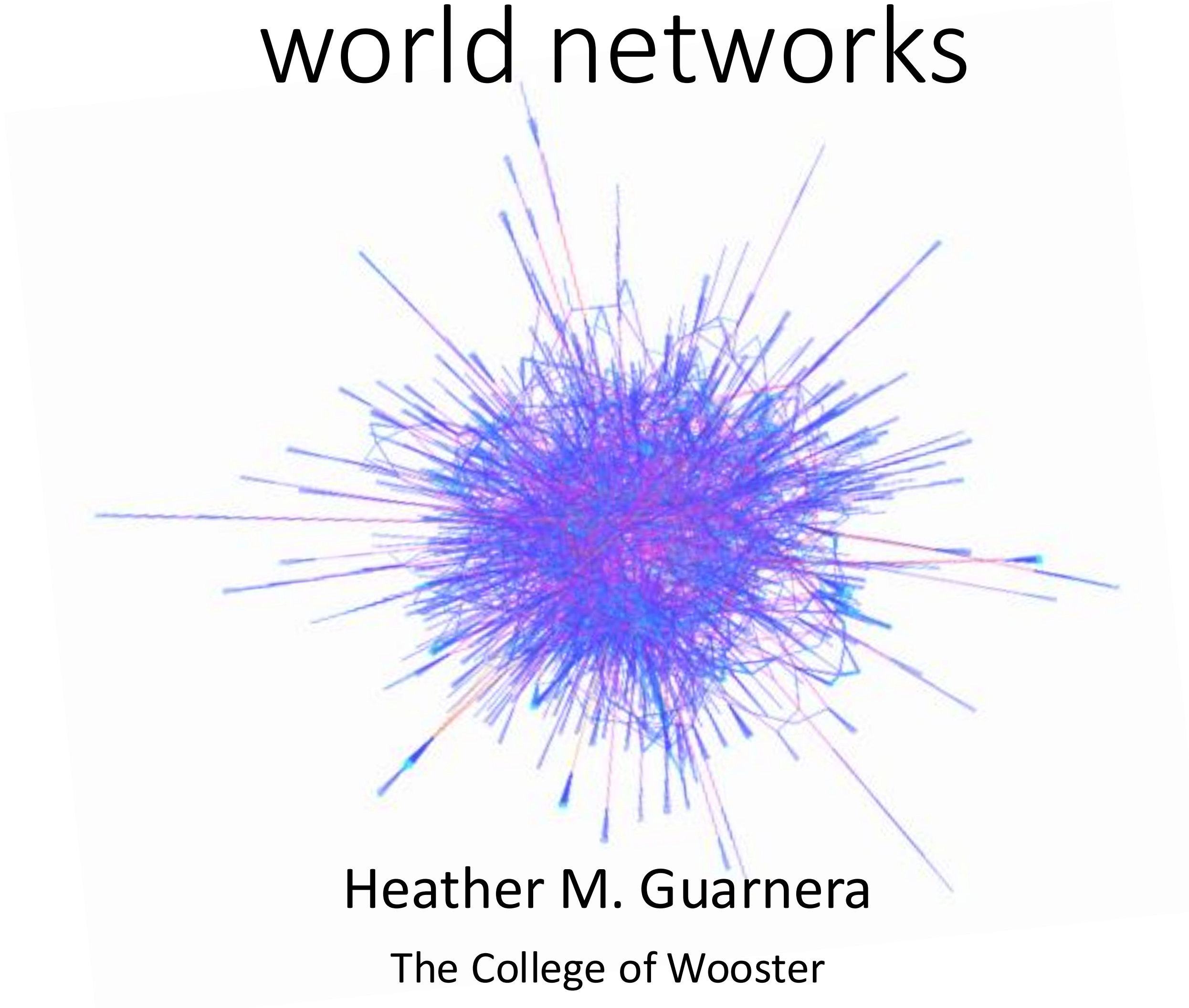
Negative curvature in real-world networks

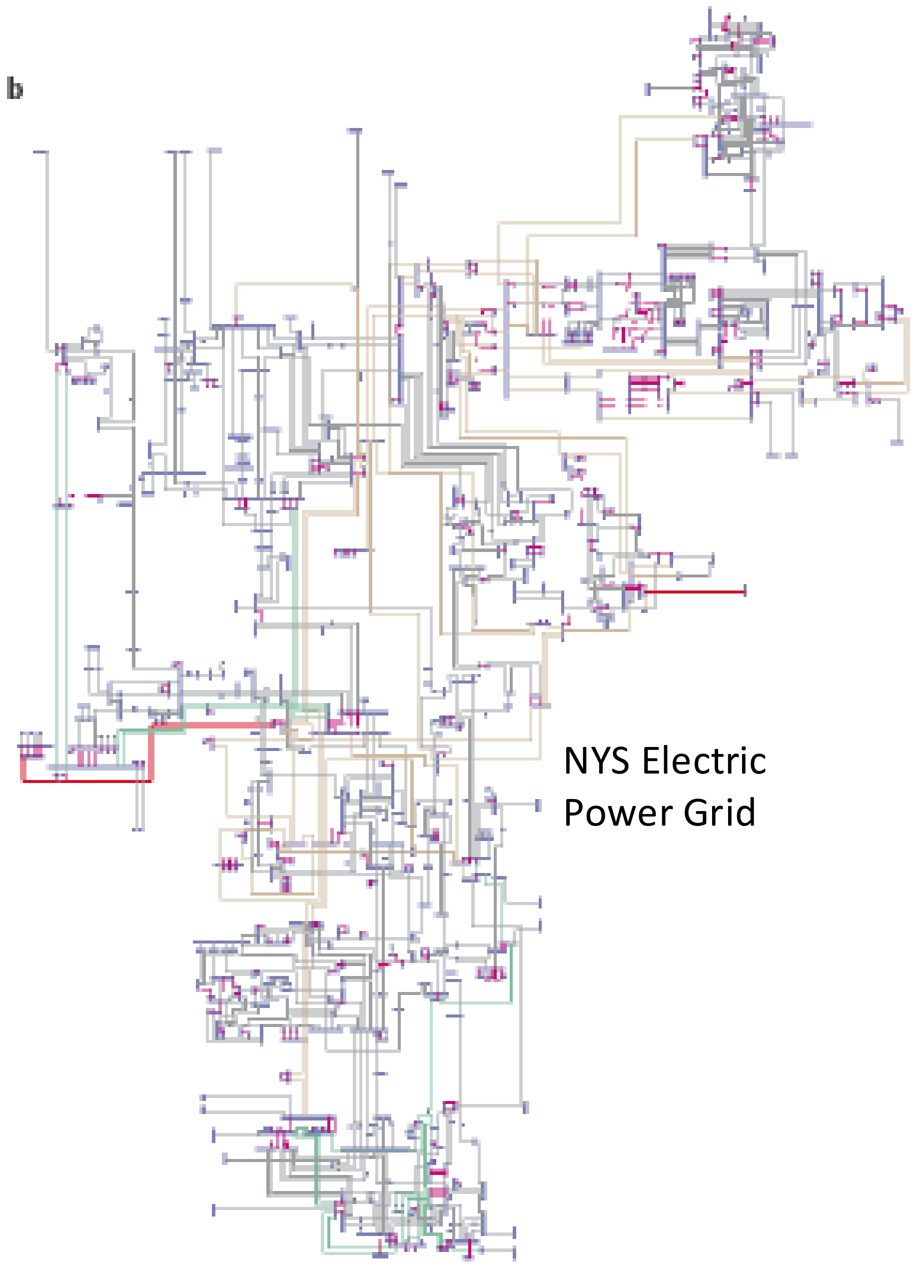


Heather M. Guarnera

The College of Wooster

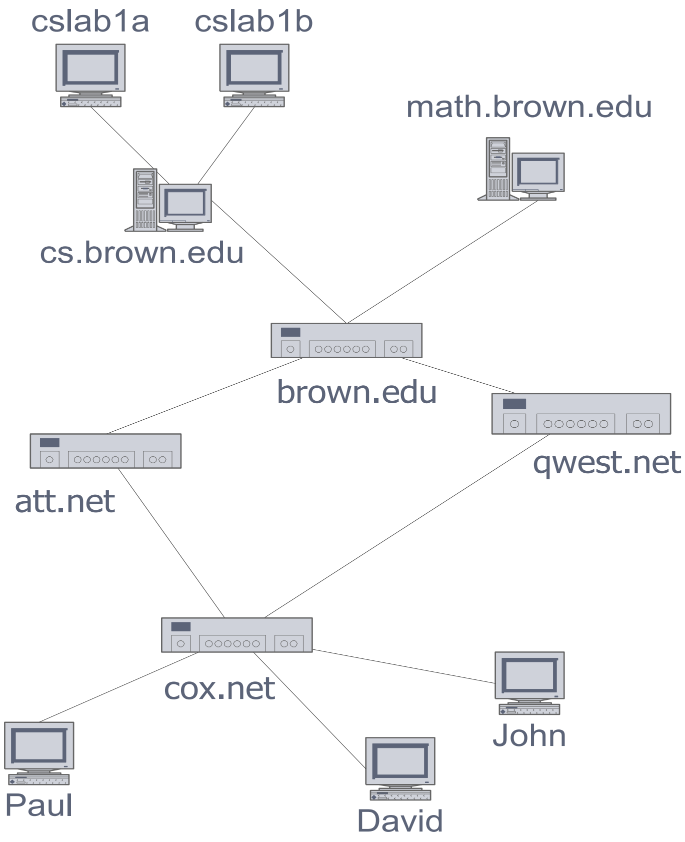
Why graph networks?

Graphs are everywhere



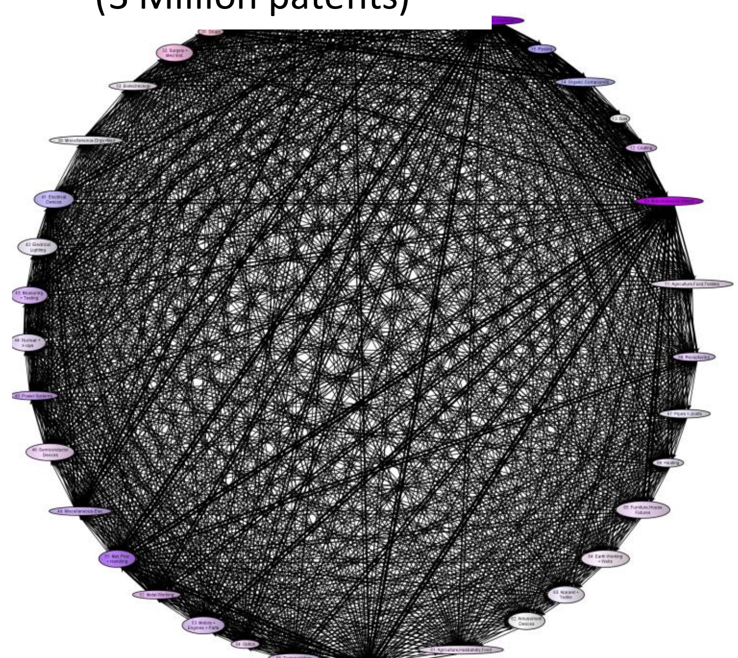
NYS Electric Power Grid

computer network

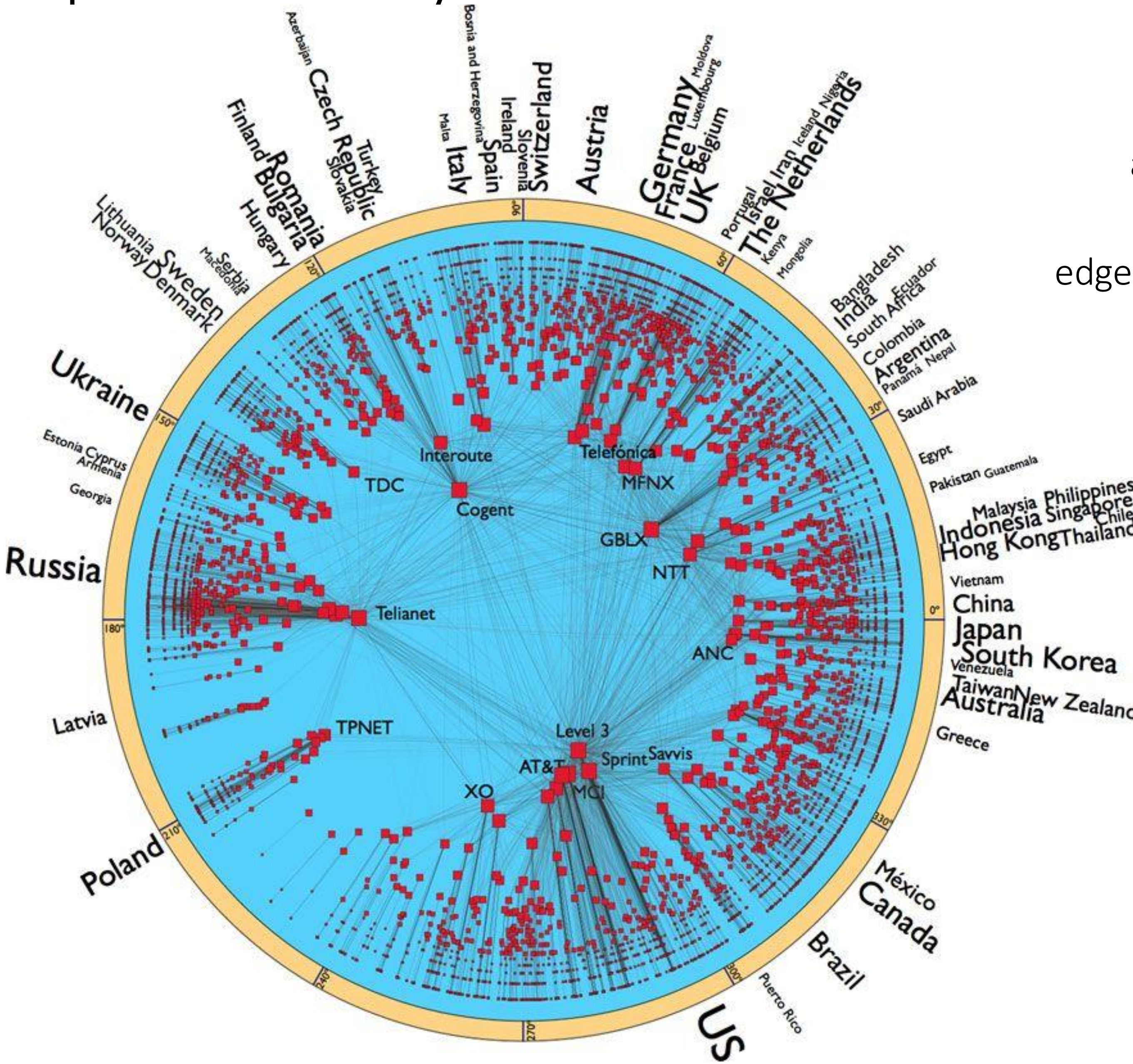


Graphs are everywhere

Utility Patent network
1972-1999
(3 Million patents)



Graphs are everywhere

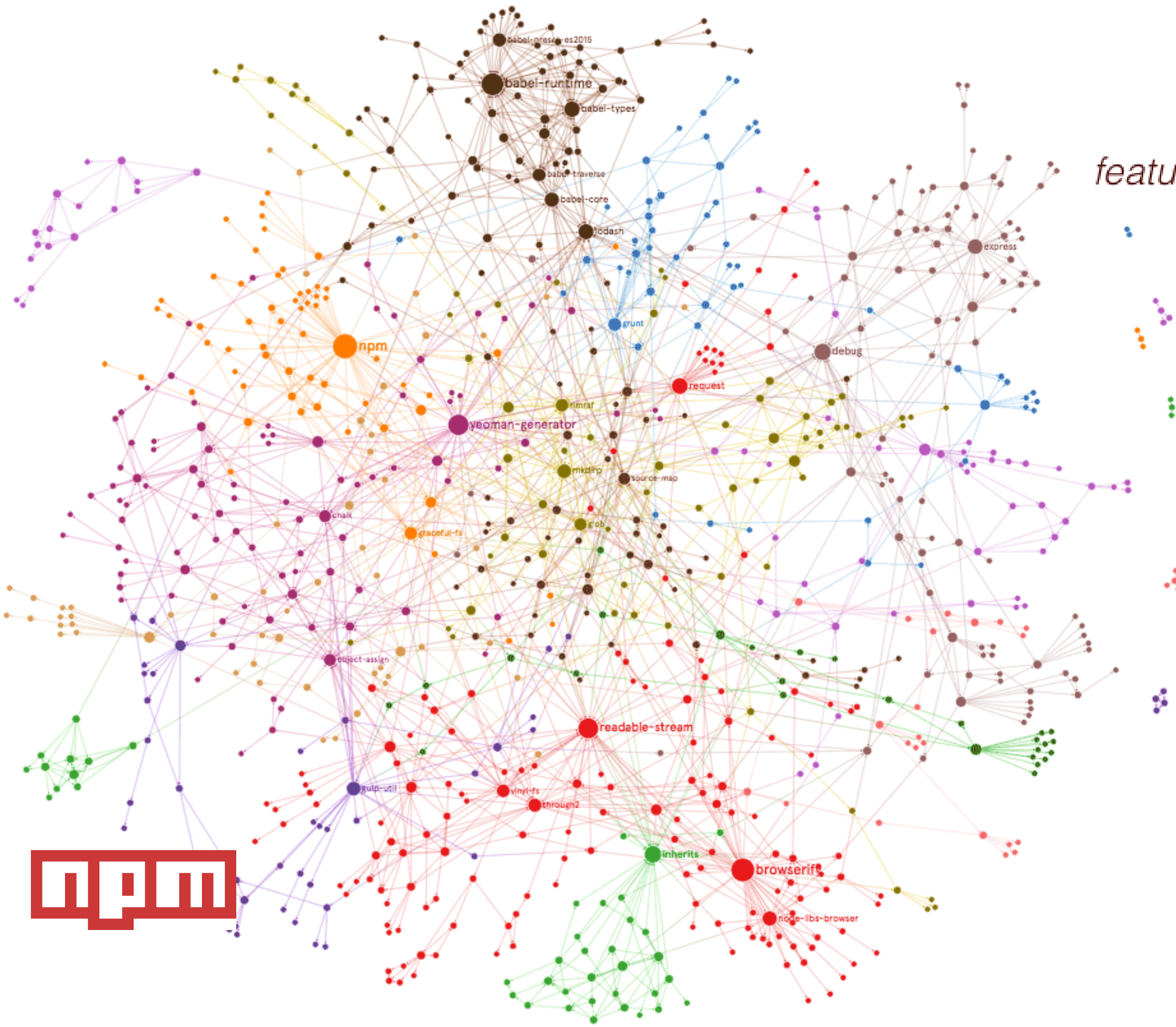


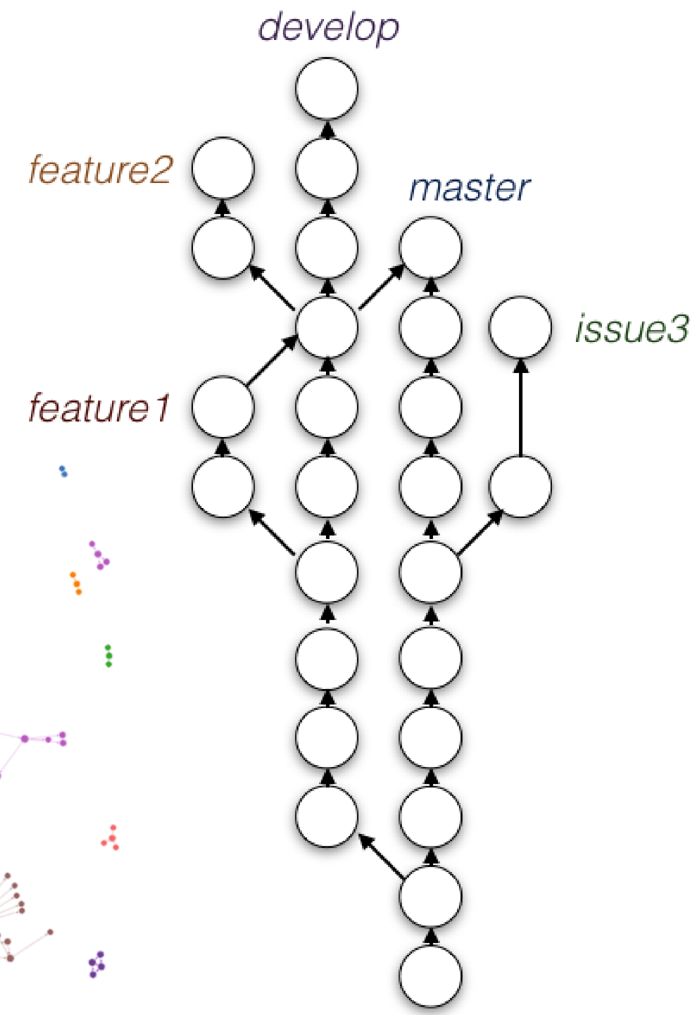
Internet (AS-level)

nodes n = 23,752 autonomous systems

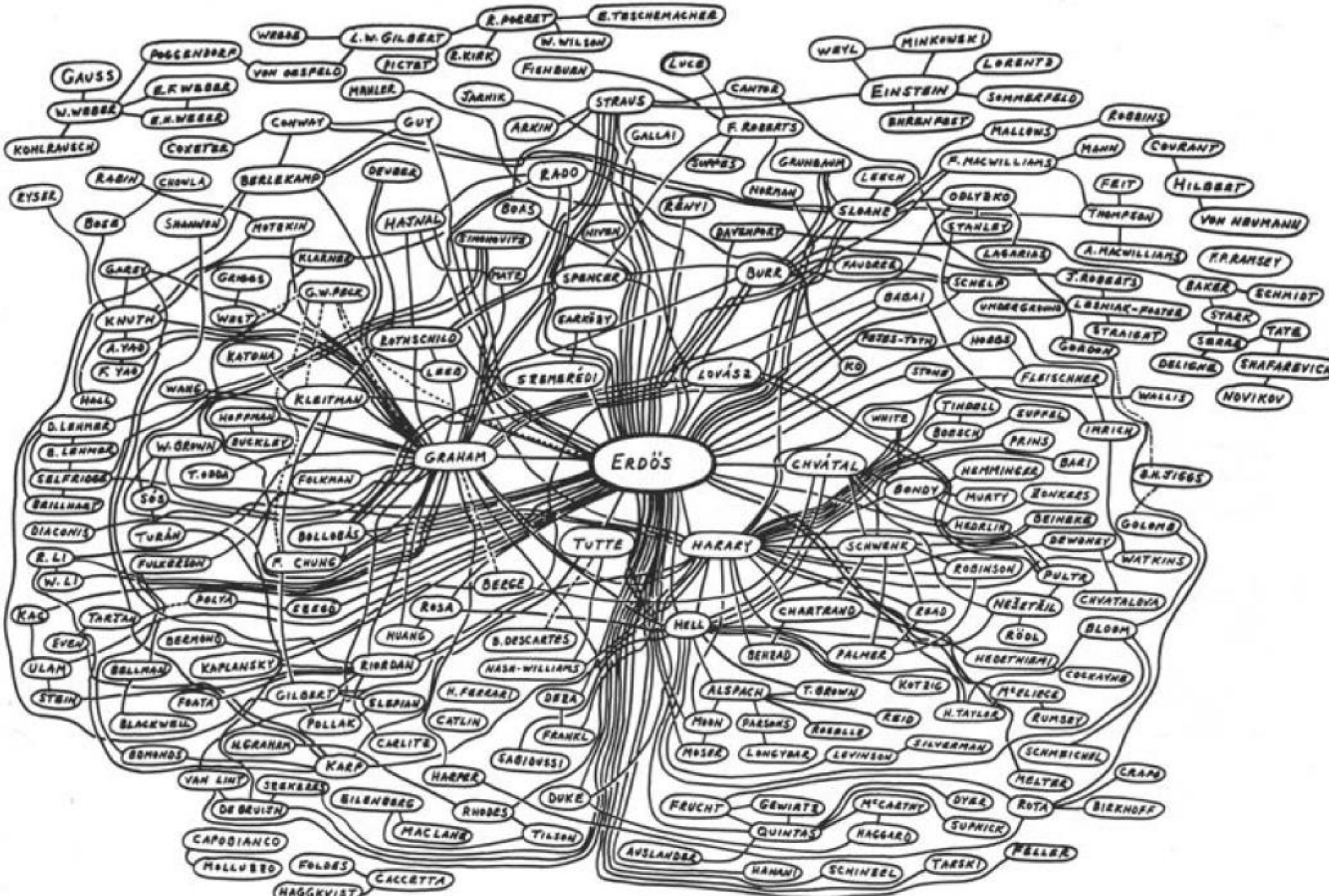
edges m = 58,416 AS links

Graphs are everywhere





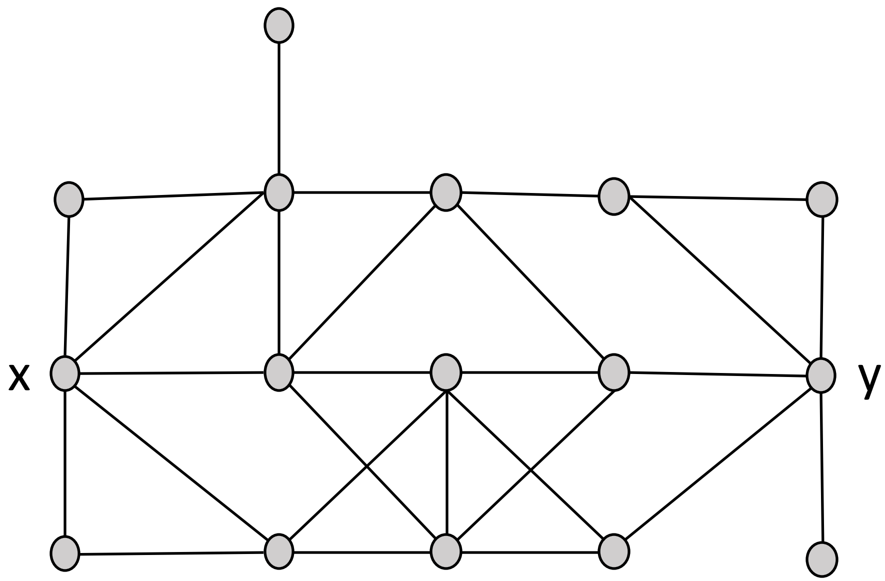
Graphs are everywhere



What is Fellow Travelers Phenomenon?

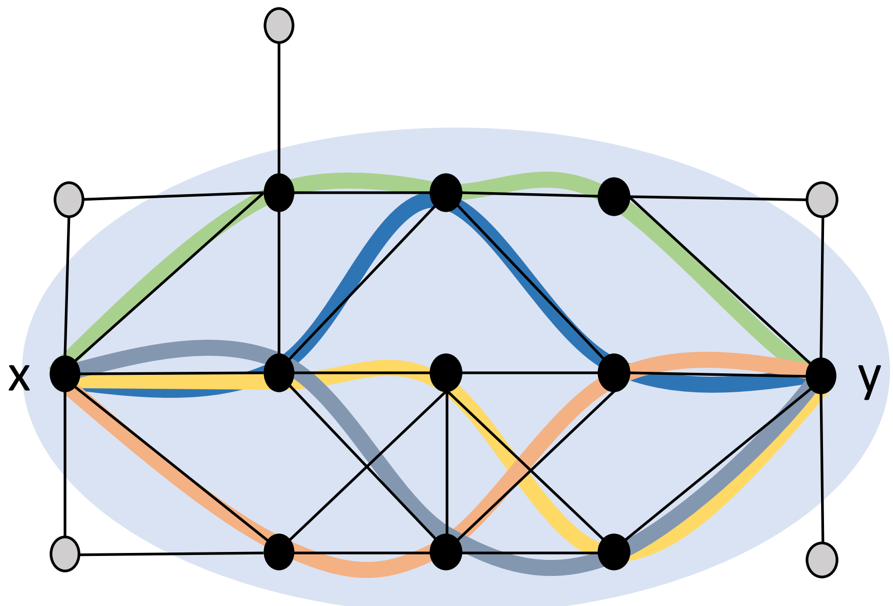
(Interval) Thinness of graphs

For any two *x,y* vertices on a graph $I(x, y) = \{z \in V : d(x, y) = d(x, z) + d(z, y)\}$ denotes the (metric) **interval**, i.e., all vertices that lay on a shortest path between x and y.



(Interval) Thinness of graphs

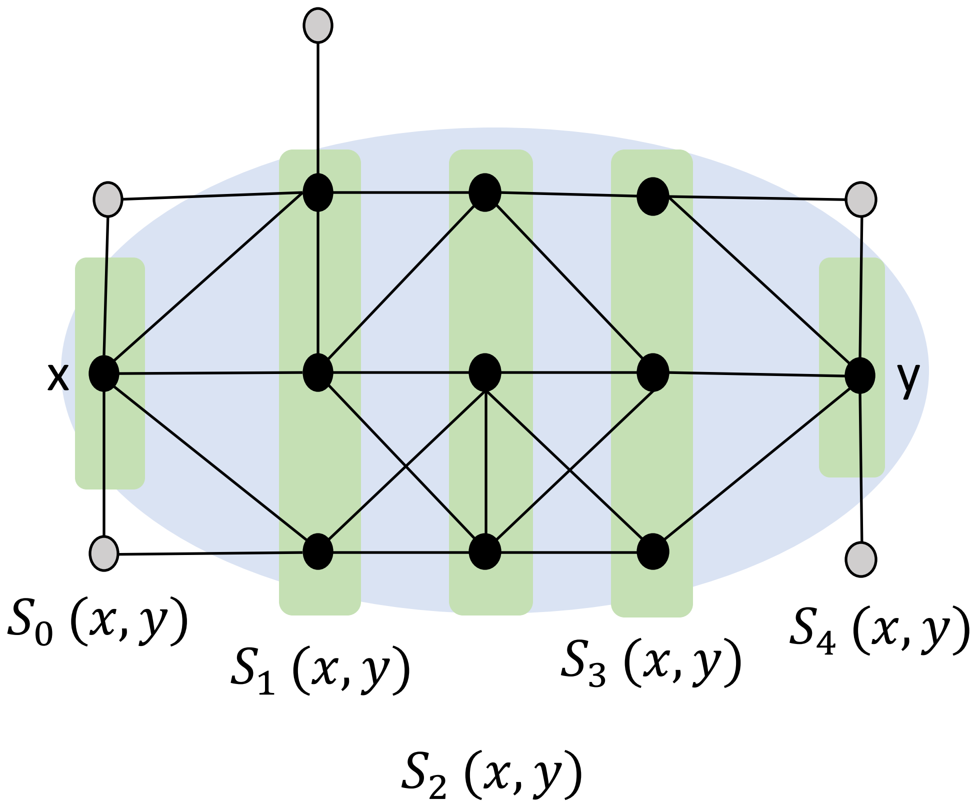
For any two *x,y* vertices on a graph $I(x,y) = \{z \in V : d(x,y) = d(x,z) + d(z,y)\}$ denotes the (metric) **interval**, i.e., all vertices that lay on a shortest path between x and y.



(Interval) Thinness of graphs

For any two *x,y* vertices on a graph $I(x,y) = \{z \in V : d(x,y) = d(x,z) + d(z,y)\}$ denotes the (metric) **interval**, i.e., all vertices that lay on a shortest path between x and y.

The set $S_p(x,y) = \{z \in I(x,y) : d(z,x) = p\}$ is called a **slice** of the interval from x to y.

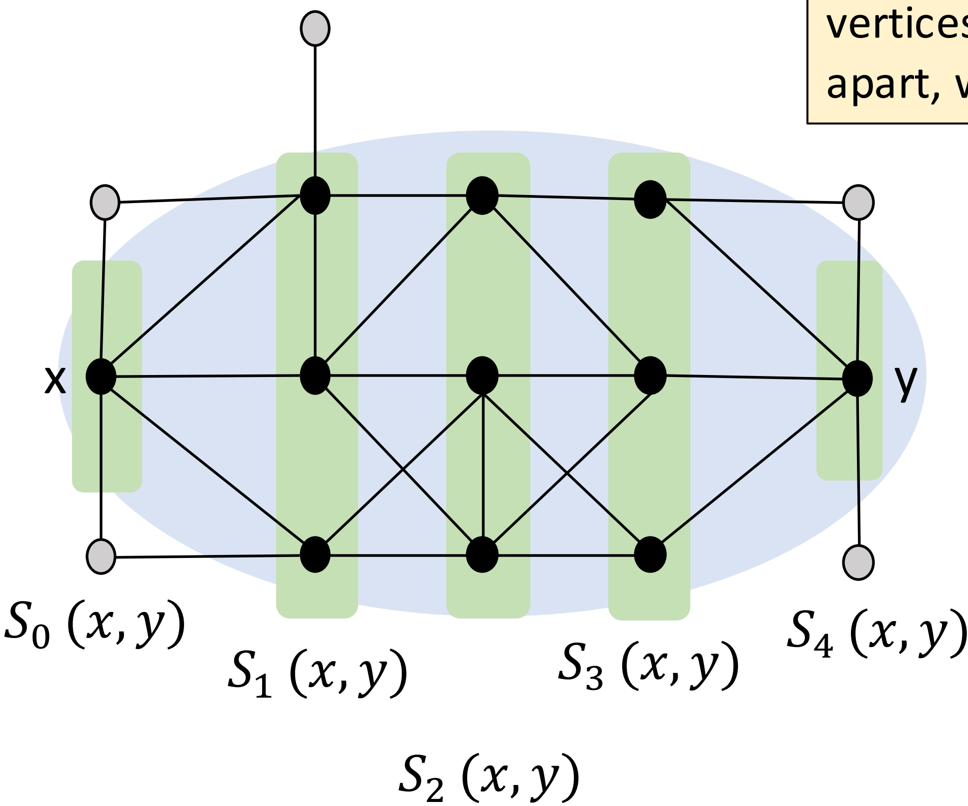


(Interval) Thinness of graphs

For any two *x,y* vertices on a graph $I(x,y) = \{z \in V : d(x,y) = d(x,z) + d(z,y)\}$ denotes the (metric) **interval**, i.e., all vertices that lay on a shortest path between x and y.

The set $S_p(x,y) = \{z \in I(x,y) : d(z,x) = p\}$ is called a **slice** of the interval from x to y.

An interval $I(x,y)$ is said to be $\boldsymbol{\kappa}$**-thin** if any two vertices u, v of the slice $S_p(x,y)$ are at most κ apart, where integer p satisfies $0 \leq p \leq d(x,y)$.



Ex: $I(x,y)$ is 2-thin.

The smallest value κ for which all intervals of G are κ-thin is the **thinness of the graph**, denoted $\boldsymbol{\kappa(G)}$.

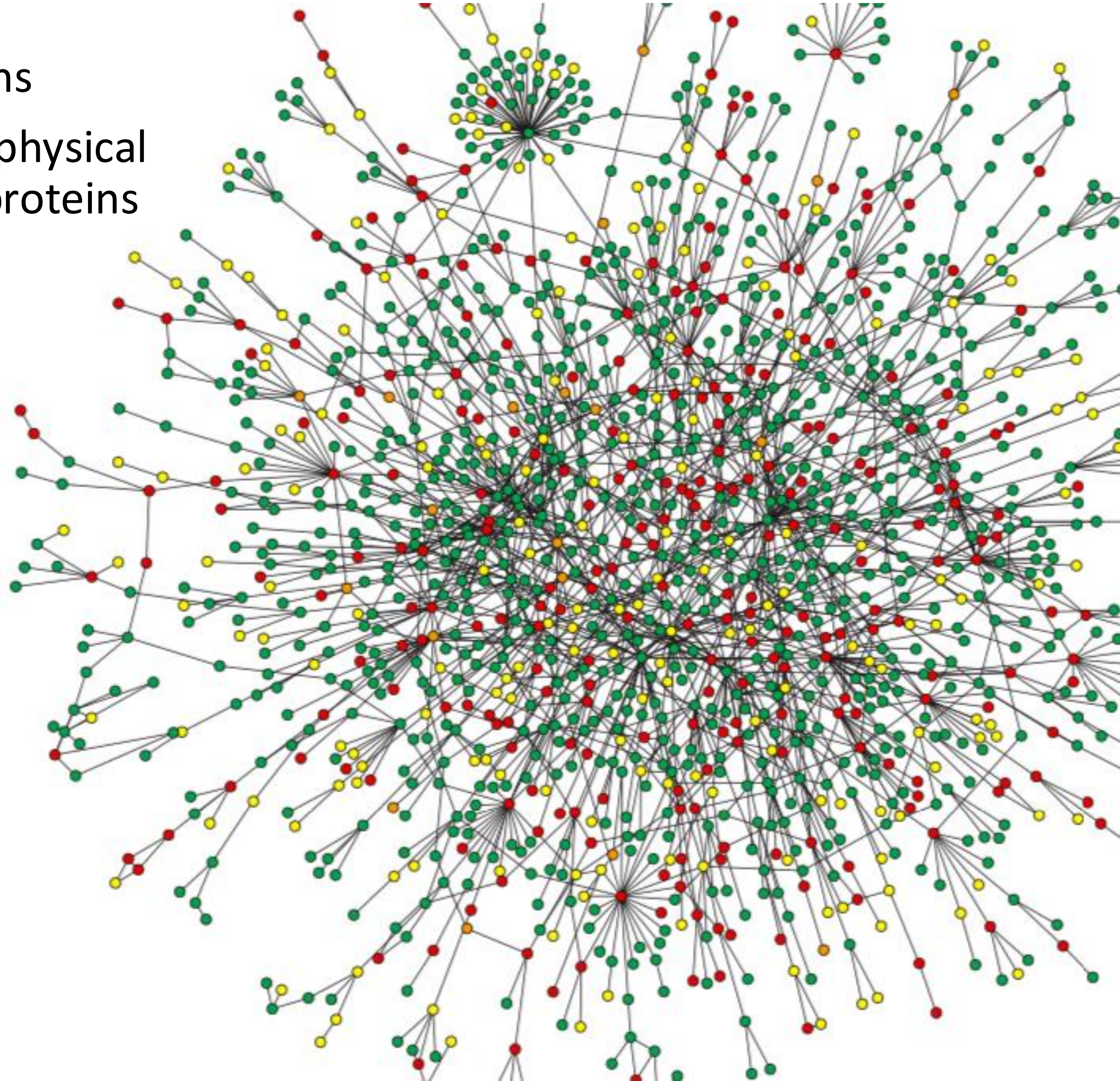
$\boldsymbol{\kappa(G)}$ is a small constant in many real-world networks!

Ex: Protein Interaction Network

nodes n = 1,870 proteins

edges m = 2240 direct physical interactions between proteins

$$\boldsymbol{\kappa(G) \leq 7}$$



Ex: Other real-world networks with small thinness

- Social networks (subset of Facebook)
  - nodes n = 293,501 users
  - edges m = 5,589,802 friendships between users

$$\boldsymbol{\kappa(G) \leq 7}$$

- Web networks (from Google)
  - nodes n = 855,802 websites
  - edges m = 4,291,352 hyperlinks connecting sites

$$\boldsymbol{\kappa(G) \leq 4}$$

- Peer-to-peer networks (Gnutella)
  - nodes n = 62,561 hosts
  - edges m = 147,878 connections between hosts

$$\boldsymbol{\kappa(G) \leq 5}$$

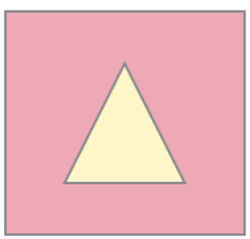
Fellow travelers phenomenon is attributed to the negative curvature of the graph

Geometric characteristics of real-world networks

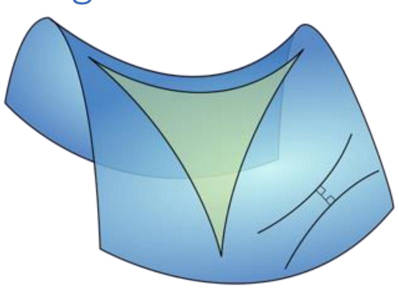
- Surge of recent empirical and theoretical work analyzes geometric characteristics

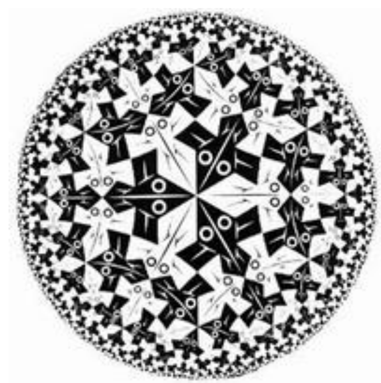
- One important property: **negative curvature**
  - causes traffic between vertices to pass through a relatively small core of the network – as if the shortest paths between them were curved inwards
  - measured in many different (somewhat equivalent) ways

Zero Curvature



Negative Curvature





Geometric characteristics of real-world networks

- Surge of recent empirical and theoretical work analyzes geometric characteristics

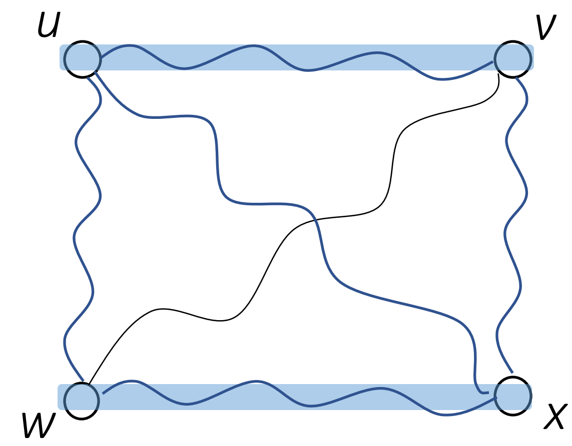
- One important property: **negative curvature**
  - causes traffic between vertices to pass through a relatively small core of the network – as if the shortest paths between them were curved inwards
  - measured in many different (somewhat equivalent) ways

- Measures of negative curvature
  - κ Interval thinness
  - τ Geodesic triangle thinness
  - δ **Gromov Hyperbolicity**
  - ς Slimness
  - ι Rooted Insize

δ-Hyperbolicity

Definition (Gromov's 4-point condition)

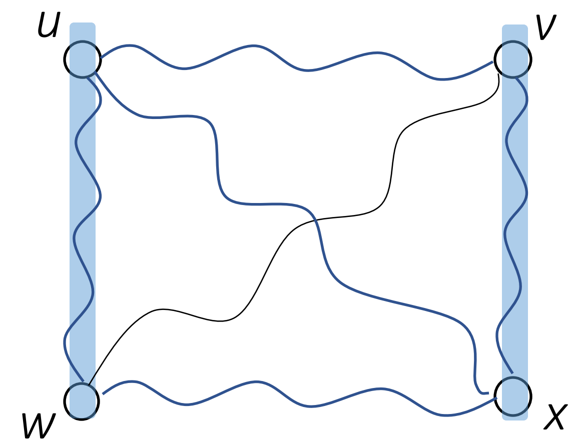
For any four points u,v,w,x, the two larger of the distance sums $d(u,v)+d(w,x)$, $d(u,w)+d(v,x)$, $d(u,x)+d(v,w)$ differ by at most $2\delta \geq 0$.



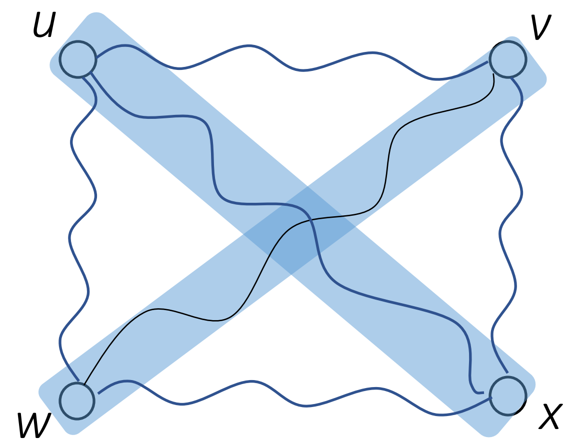
δ-Hyperbolicity

Definition (Gromov's 4-point condition)

For any four points *u,v,w,x,* the two larger of the distance sums *d(u,v)+d(w,x), d(u,w)+d(v,x), d(u,x)+d(v,w)* differ by at most $2\delta \geq 0$.



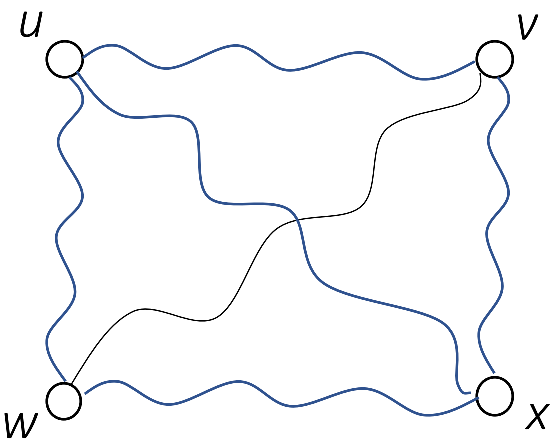
δ-Hyperbolicity



Definition (Gromov's 4-point condition)

For any four points *u,v,w,x,* the two larger of the distance sums *d(u,v)+d(w,x), d(u,w)+d(v,x), d(u,x)+d(v,w)* differ by at most $2\delta \geq 0$.

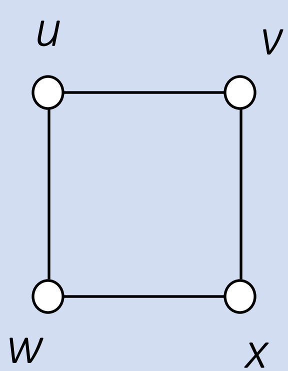
δ-Hyperbolicity



Definition (Gromov's 4-point condition)

For any four points *u,v,w,x,* the two larger of the distance sums *d(u,v)+d(w,x), d(u,w)+d(v,x), d(u,x)+d(v,w)* differ by at most $2\delta \geq 0$.

Example:



d(u,v) + d(w,x) = 2

d(u,w) + d(v,x) = 2

d(u,x) + d(v,w) = 4

So, $\delta = \frac{4-2}{2} = 1$

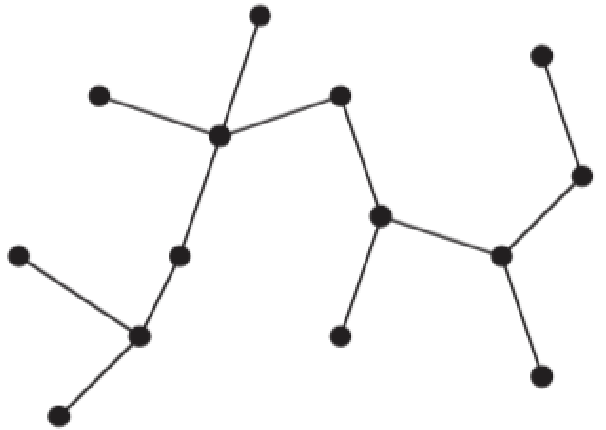
Take any quadruple of vertices and these 3 distances sums.

$2\delta \geq$ LargestSum - MiddleSum

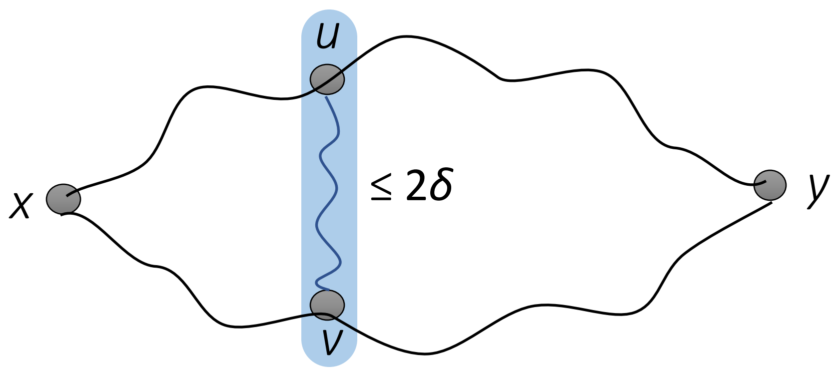
Relation of interval thinness to hyperbolicity

δ-Hyperbolicity measures how close (locally) a metric space is to a tree from a metric point of view; the smaller the value indicate

- is metrically closer to a tree ($\delta=0$ in a tree)
- has global negative curvature

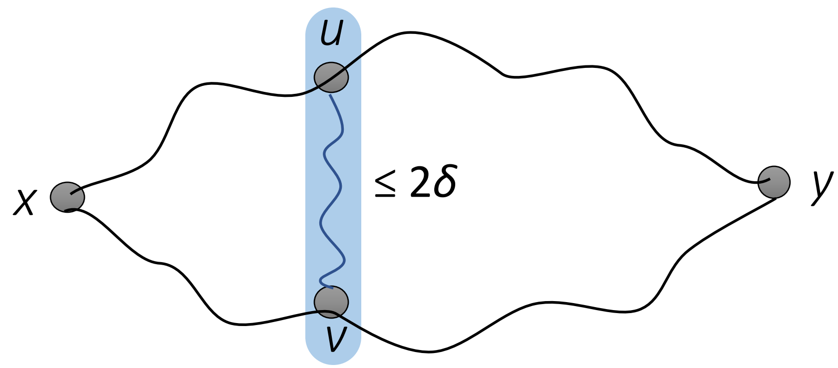


Lemma (**Fellow travelers property**): For any graph G, $\kappa(G) \leq 2\delta(G)$.



Relation of interval thinness to hyperbolicity

Lemma (**Fellow travelers property**): For any graph G, κ(G) ≤ 2δ(G).



Proof:

Let $x, y \in V$, and let u, v belong to the same slice of the interval $I(x, y)$.
Consider the 3 distance sums between these 4 vertices.

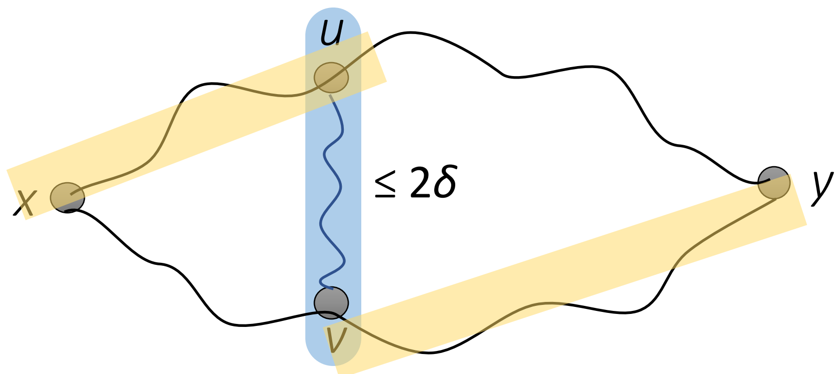
$$d(x,u) + d(v,y)$$

$$d(x,v) + d(u,y)$$

$$d(x,y) + d(u,v)$$

Relation of interval thinness to hyperbolicity

Lemma (**Fellow travelers property**): For any graph G, κ(G) ≤ 2δ(G).



Proof:

Let $x, y \in V$, and let u, v belong to the same slice of the interval $I(x, y)$.
Consider the 3 distance sums between these 4 vertices.

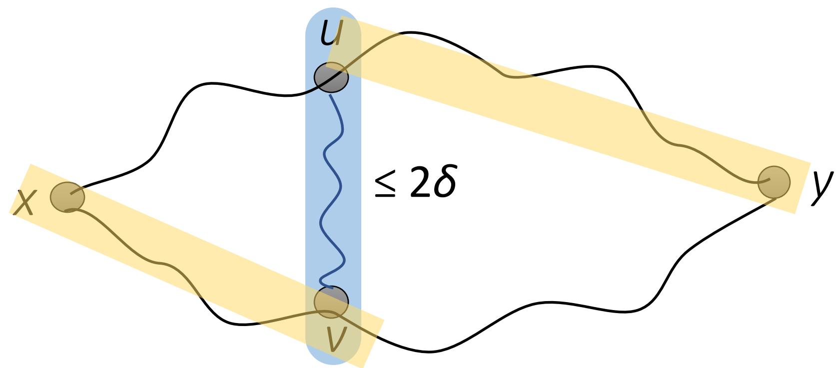
$$d(x,u) + d(v,y) \quad = d(x,y)$$

$$d(x,v) + d(u,y)$$

$$d(x,y) + d(u,v)$$

Relation of interval thinness to hyperbolicity

Lemma (**Fellow travelers property**): For any graph G, κ(G) ≤ 2δ(G).



Proof:

Let $x, y \in V$, and let u, v belong to the same slice of the interval $I(x, y)$.
Consider the 3 distance sums between these 4 vertices.

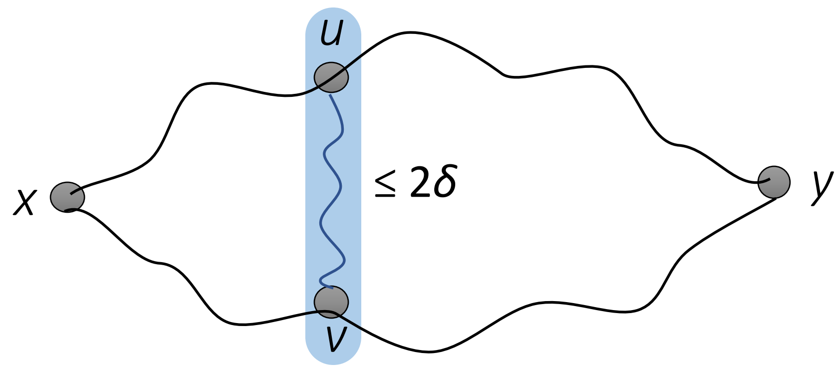
$$d(x,u) + d(v,y) \quad = d(x,y)$$

$$d(x,v) + d(u,y) \quad = d(x,y)$$

$$d(x,y) + d(u,v)$$

Relation of interval thinness to hyperbolicity

Lemma (**Fellow travelers property**): For any graph G, $\kappa(G) \leq 2\delta(G)$.



Proof:

Let $x, y \in V$, and let u, v belong to the same slice of the interval $I(x, y)$.
Consider the 3 distance sums between these 4 vertices.

$$d(x,u) + d(v,y) \quad = d(x,y)$$

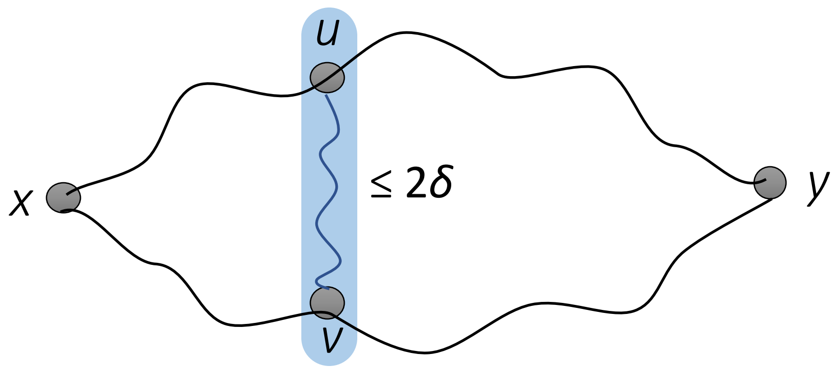
$$d(x,v) + d(u,y) \quad = d(x,y)$$

$$d(x,y) + d(u,v) \longleftarrow \text{Largest Sum}$$

From definition of hyperbolicity, $2\delta \geq d(x,y) + d(u,v) - d(x,y) = d(u,v)$.

Relation of interval thinness to hyperbolicity

Lemma (**Fellow travelers property**): For any graph G, κ(G) ≤ 2δ(G).



Theorem [1]: For every **Helly** graph G, κ(G) ≤ 2δ(G) ≤ κ(G)+1.

Open question: What other types of graphs behave in this way?

[1] F. Dragan, **H. Guarnera**, "Obstructions to a small hyperbolicity in Helly graphs", Discrete Mathematics, 342(2):326 – 338, 2019.

How can this geometric information be applied?

Parameterized complexity/approximation factor

- **Goal**: create algorithms which solve problems utilizing these geometric properties

- Example: Consider δ hyperbolicity, which is known to be small in many real-world networks.
  - Solve a problem in $O(f(\delta)\ m)$ time
  - Compute a $f(\delta)$ approximation

- Some problems this has been applied to:
  - Covering/packing problems
  - Computing the diameter/radius
  - Facility location problems
  - Network analysis
  - Vertex pursuit games on graphs
  - Traveling salesman problem

Parameterized complexity/approximation factor

- **Goal**: create algorithms which solve problems utilizing these geometric properties

- Example: Consider δ hyperbolicity, which is known to be small in many real-world networks.
  - Solve a problem in $O(f(\delta)\, m)$ time
  - Compute a $f(\delta)$ approximation

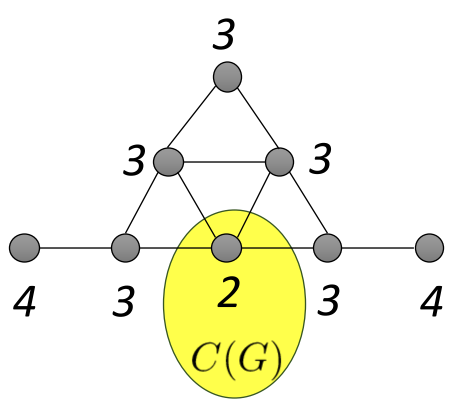
- Some problems this has been applied to:
  - Covering/packing problems
  - Computing the diameter/radius
  - Facility location problems
  - Network analysis
  - Vertex pursuit games on graphs
  - Traveling salesman problem

1. F. Dragan and **H. Guarnera**. Helly-gap of a graph and vertex eccentricities. Theoretical Computer Science, 867:68-84, 2021.
2. F. Dragan and **H. Guarnera**. Eccentricity function in distance-hereditary graphs. Theoretical Computer Science, 833: 26-40, 2020.
3. F. Dragan and **H. Guarnera**. Eccentricity terrain of δ-hyperbolic graphs. Journal of Computer and System Sciences, 112: 50-56, 2020.
4. F. Dragan, G. Ducoffe, **H. Guarnera**. Fast deterministic algorithms for computing all eccentricities in (hyperbolic) Helly graphs, the 17th Algorithms and Data Structures Symposium (WADS'21), 2021.
5. Mohammed, F. Dragan, **H. Guarnera**. Fellow Travelers Phenomenon Present in Real-World Networks, Complex Networks & Their Applications, 2022.

Example: eccentricity function and centers

The eccentricity *e(x)* of a vertex *x* is the distance to a furthest *u* vertex to *x*

$$e(x) = \max_{u \in V} d(x, u)$$



The minimum and maximum eccentricities are called the radius *rad(G)* and diameter *diam(G)* of the graph, respectively

The center of a graph *C(G)* is the set of vertices with minimum eccentricity

$$C(G) = \{v \in V : e(v) = rad(G)\}$$

Applications:

- Measure the importance of a node (centrality indices)
- Facility location problems
- Detecting small-world networks (degrees of freedom)

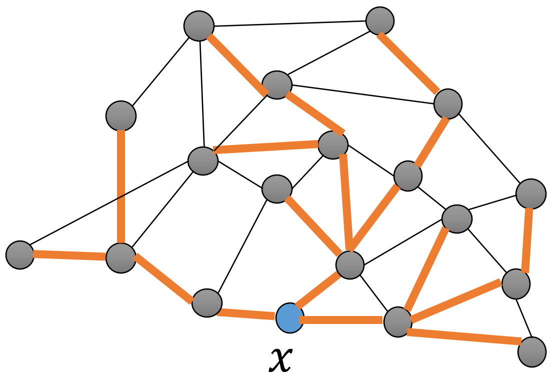
Computing vertex eccentricities straightforwardly.

The eccentricity *e(x)* of a vertex *x* is the distance to a furthest *u* vertex to *x*

$$e(x) = \max_{u \in V} d(x, u)$$

Take a connected graph with n vertices and m edges.

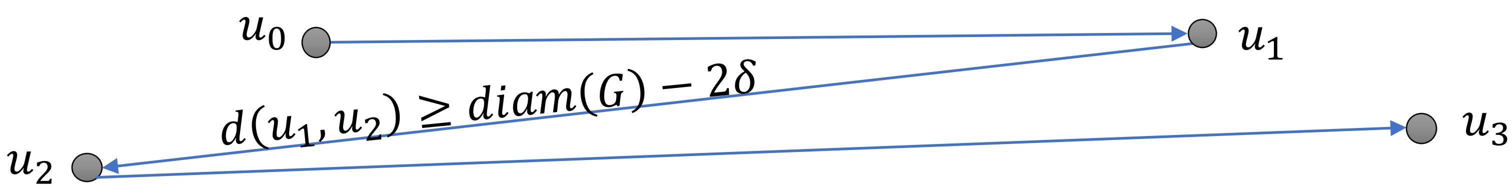
- A single Breadth-First Search (BFS) from a vertex x
  - runs in *O(m)* time
  - yields *e(x)*
- Call BFS for each of the n vertices
- Total *O(nm)* runtime



This is prohibitively expensive on many real-world networks, as they are huge!

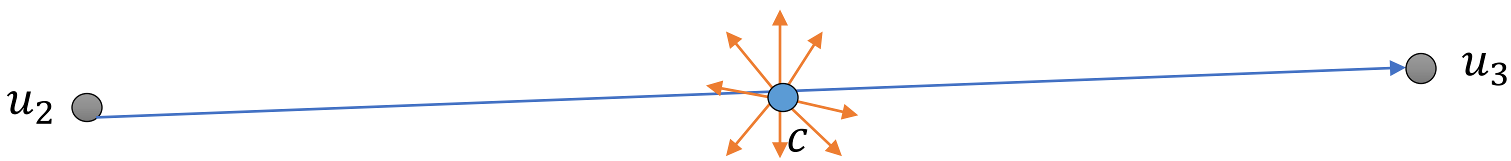
Efficient eccentricity approximation via eccentricity approximating spanning tree

- Find a long path in $O(m)$ time

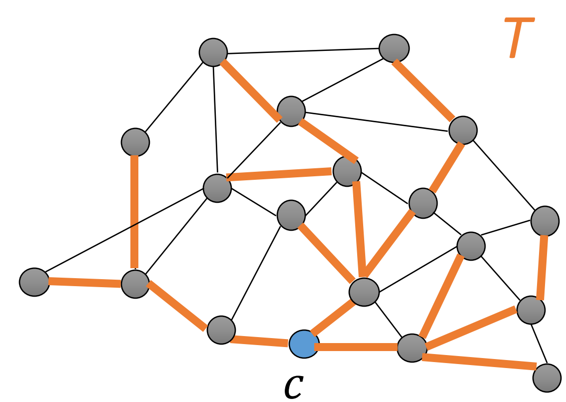


Efficient eccentricity approximation via eccentricity approximating spanning tree

- Find a long path in $O(m)$ time



- Run breadth-first search (BFS) from the middle vertex *c* between $u_2 u_3$
- We show $e_T(v) \leq e_G(v) \leq e_T(v) + 6\delta$



Theorem [2]: There is a 6δ approximation of all eccentricities in total $O(m)$ time

[2] F. Dragan and **H. Guarnera**. Eccentricity terrain of δ-hyperbolic graphs. Journal of Computer and System Sciences, 112: 50-56, 2020.

Conclusion

- Many real world networks exhibit the fellow travelers property
  - Biological networks
  - Communication networks
  - Social networks
  - Software ecosystems

- We can take advantage of this nice geometric property to solve problems faster on these networks
  - Ex: computing vertex eccentricities

Conclusion and future work

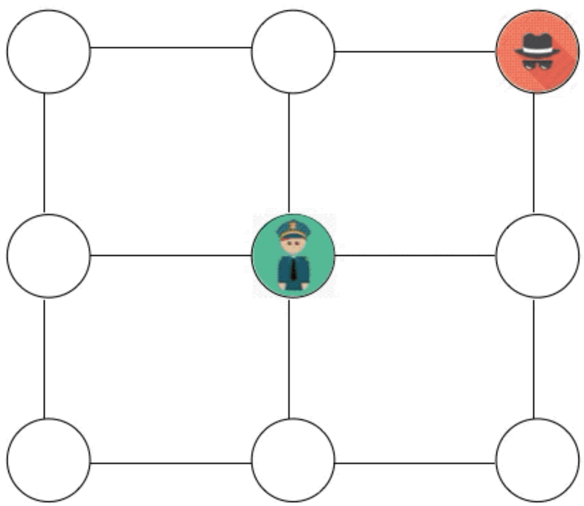
- Many real world networks exhibit the fellow travelers property
  - Biological networks
  - Communication networks
  - Social networks
  - Software ecosystems
  - **What else?**

- We can take advantage of this nice geometric property to solve problems faster on these networks
  - Ex: computing vertex eccentricities
  - **What else? Ex: vertex pursuit games**

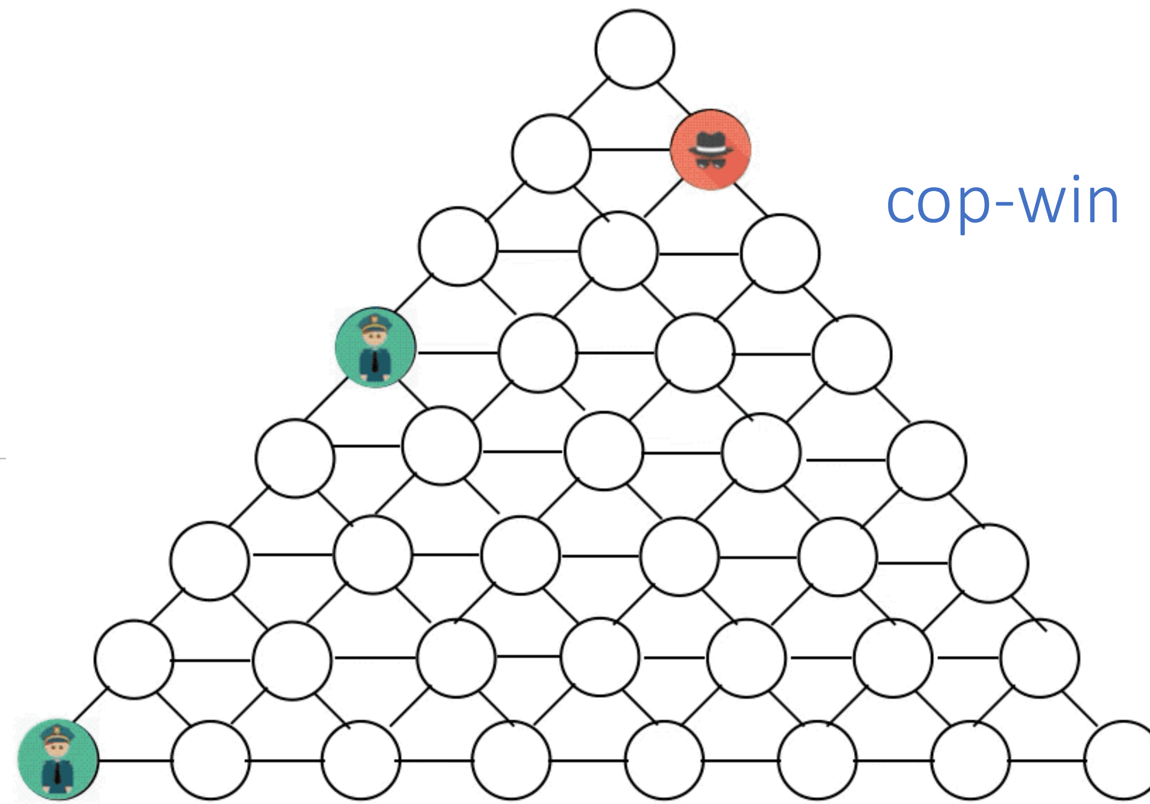
- How does interval thinness relate to other geometric measures of negative curvature?

- What other problems can be solved better with interval thinness, compared to other measures?

Vertex pursuit games on graphs: cops vs. robbers

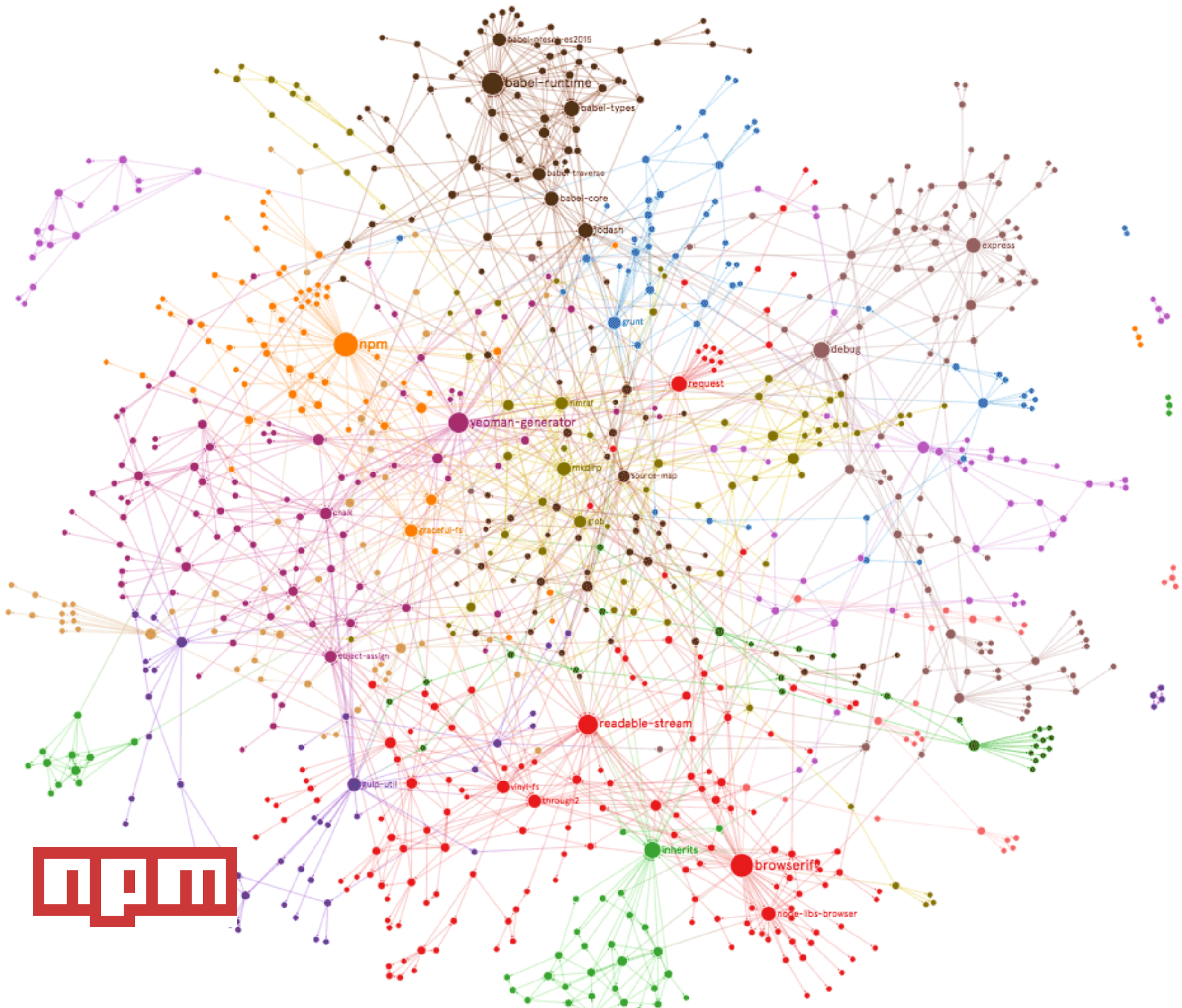


robber-win



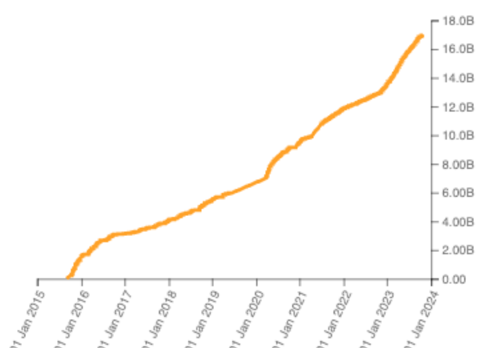
cop-win

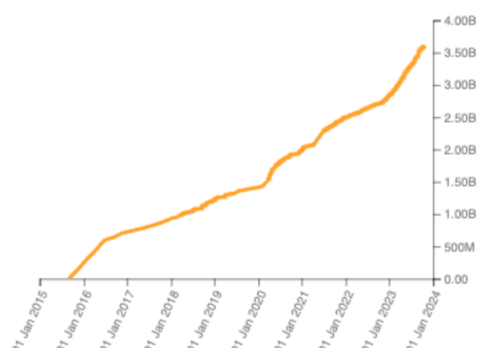
Network analysis: software ecosystems

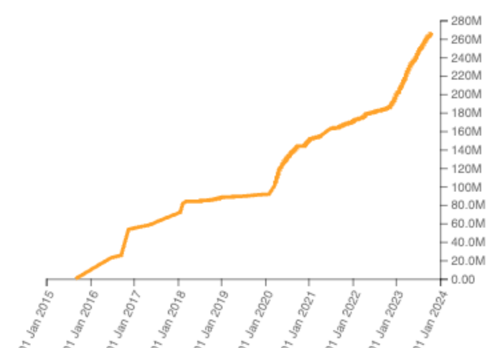


Software Heritage Project

Source files	Commits	Projects
17,001,055,063	3,605,964,127	266,259,421







Directories	Authors	Releases
13,682,190,925	66,817,476	77,841,991

Thank you! Questions?