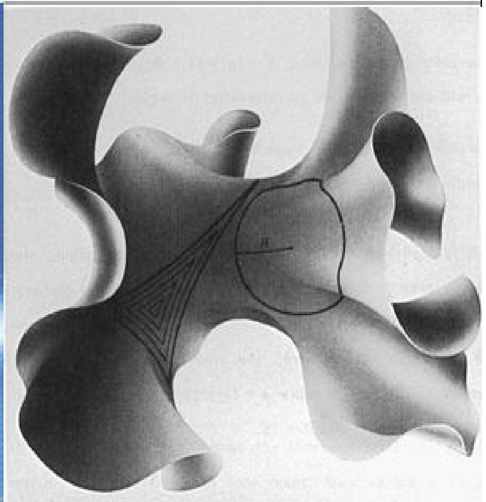
Alcatel·Lucent
Bell Labs

Large-Scale Curvature of Networks:

And Implications for Network Management and Security



Iraj Saniee

(Joint work with Onuttom Narayan, UCSC & funded by AFOSR)

Math & Computing Lab, Bell Labs, ALU

600 Mountain Ave., Murray Hill, NJ 07974

Presentation at NIST, Nov. 9th, 2010

Understanding Large-Scale Networks

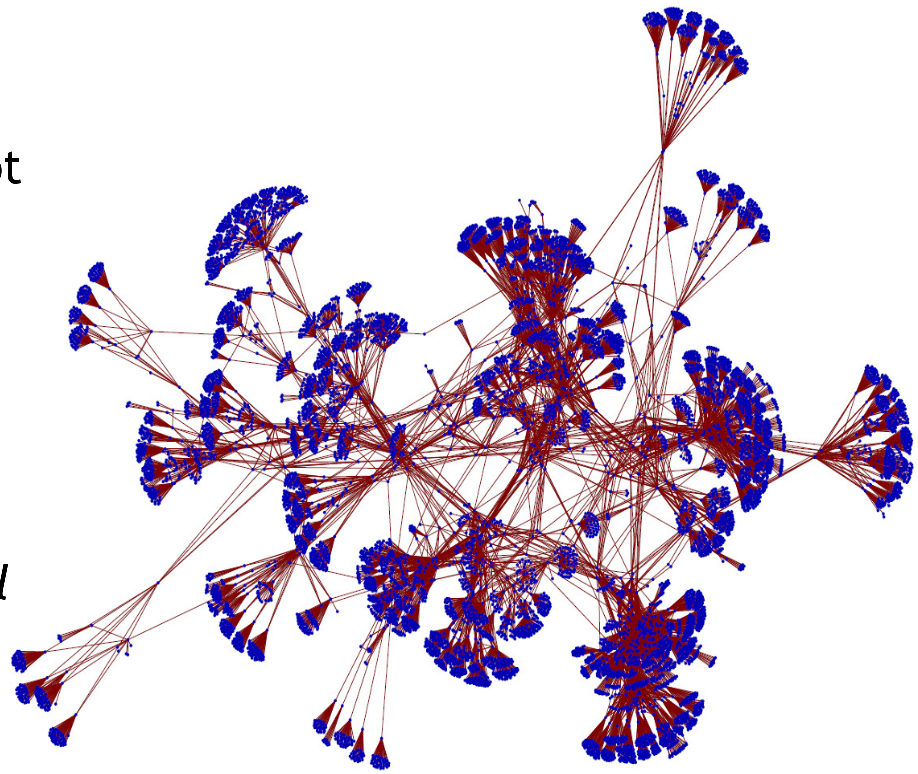
- Hard to visualize due to scale

- Unclear what is essential and what is not for overall *performance*, *reliability* and *security*

- Much of the existing work on "complex networks" focuses on *local* measures such as degree distribution, clustering coefficients, etc. at the expense of *global* properties

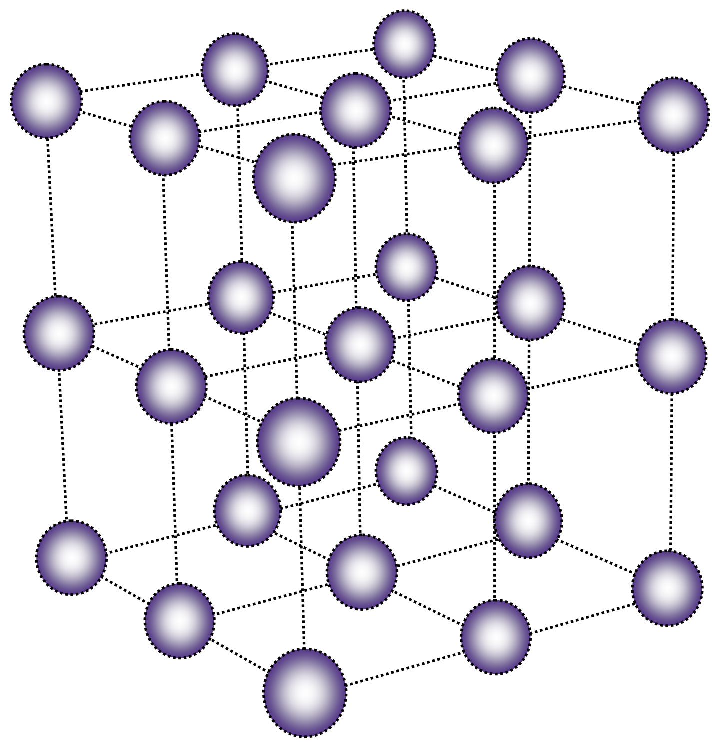
- Need more fundamental ways to "summarize" critical network information

- A promising direction is to look at key geometric characteristics of networks: ***dimension*** and ***curvature***



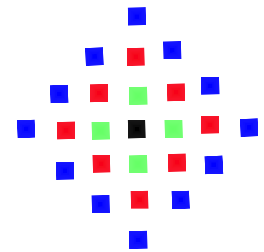
Rocketfuel dataset 7018
10152 nodes, 28638 links, diameter 12

Dimension -- Degrees of Freedom

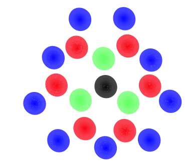


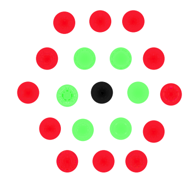
Dimension of a Lattice & Average Shortest Path Lengths

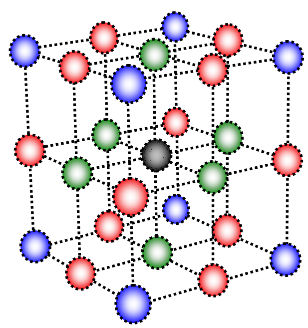
- How fast does a "typical ball" grow? Look at circumference or volume as a function of "radius"



Circumference of Configuration *(dimension D, degree d)*	*1-hop away*	*2-hops away*	*3-hops away*	*4-hops away*
Square (D=2,d=4)	$4=4*1^1$	$8=4*2^1$	$12=4*3^1$	$16=4*4^1$
Hexagon (D=2,d=3)	$3=3*1^1$	$6=3*2^1$	$9=3*3^1$	$12=3*4^1$
Triangle (D=2,d=6)	$6=6*1^1$	$12=6*2^1$	$18=6*3^1$	$24=6*4^1$
Cube (D=3,d=6)	$6=6*1^2$	$16=4*2^2$	$36=4*3^2$	$64=4*4^2$
General (D,d)	$d*1^{(D-1)}$	$d*2^{(D-1)}$	$d*3^{(D-1)}$	$d*4^{(D-1)}$





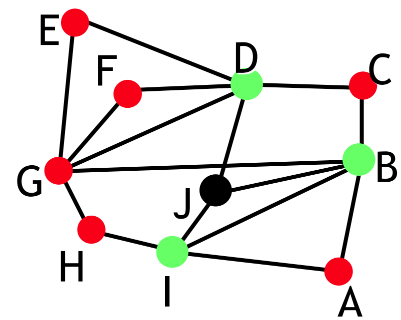


Average length of a shortest path <h> of a grid in dimension D

$$\approx (D/D+1)(DN/d)^{1/D} \approx O(N^{1/D})$$

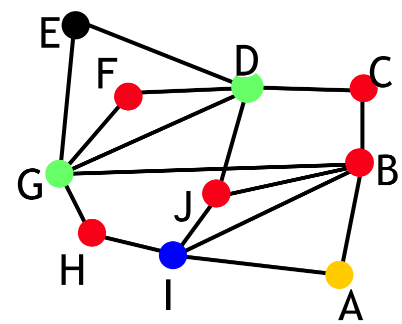
Dimension of a *Network* & Its Average Shortest Path Length

Measure the number of neighbors of a node X *h* hops away. How does this number scale with *h*? If roughly like $h^{\Delta-1}$ then we say Δ *is the dimension of the graph in the neighborhood of X*.

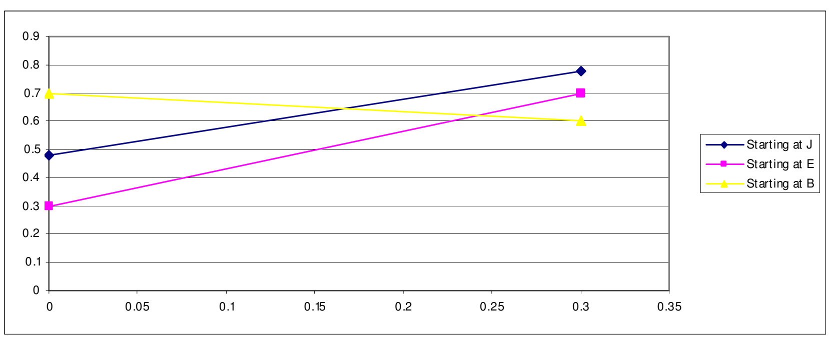


Node	*1-hop away*	*2-hops away*	*3-hops away*	*4-hops away*
J (NSF)	3	6		
E (NSF)	2	5	1	1
B (NSF)	5	4		
A (lattice)	3	3	2	
B (lattice)	3	3	2	

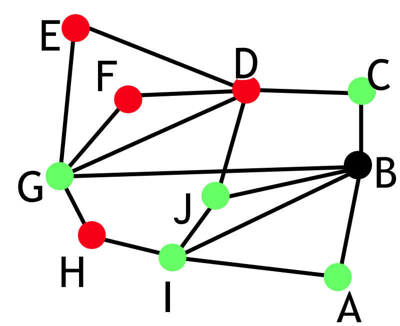
- Start node
- 1 hop away
- 2 hops away
- 3 hops away
- 4 hops away



$$v(r) \sim r^D \Rightarrow \log(v(r)) \sim D^*\log(r)$$



$D_{small\text{-}graph} \approx 1.7$



Data Source - Rocketfuel (Washington U, NSF 2002-05)

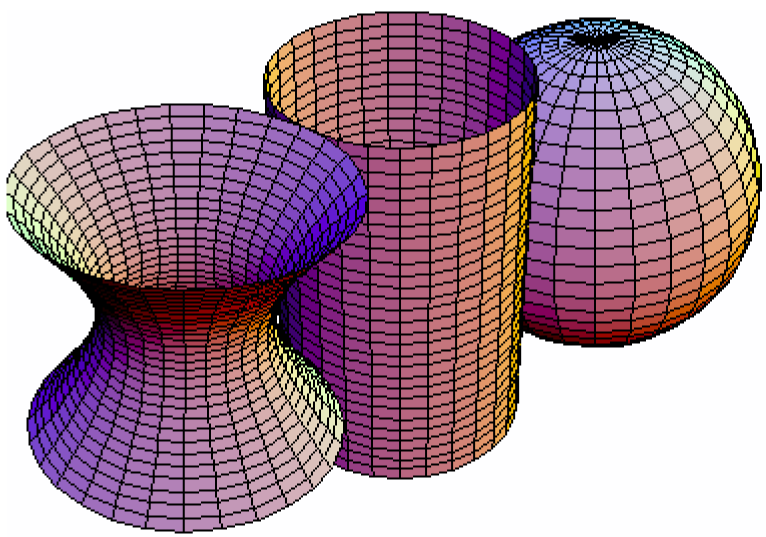
Look at scaling of the average shortest path length <*h*>

- In 2-dim grid, <*h*> $\sim\surd N$ (or $\sim N^{1/D}$ in D-dimensional grid)
- Look at "Rocketfuel" data, [Washington University researchers' detailed connectivity data from various ISPs 2002-2003]
- <*h*> does not scale like $\surd N$ or $N^{1/D}$ but are more like *log(N)* -- *"Small World"* like

=> *RF networks do not appear to be grid-like (or flat) nor do they exhibit characteristics of finite dimensions*

Network ID	Network Name	Size #node - #links	Average Shortest Path Length
1221	Telstra (Australia)	2998 - 7612	5.53
1239	Sprintlink (US)	8341 - 28050	5.18
1755	EBONE (US)	605 - 2070	6.0
2914	Verio (US)	3045 - 10726	6.0
3257	Tiscali (EU)	855 - 2346	5.3
3356	Level 3 (US)	3447 - 18780	5.0
3967	Exodus (US)	895 - 4140	5.9
4755	VSNL (India)	121 - 456	3.2
6461	Abovenet (US)	2720 - 7648	5.7
7018	AT&T (US)	10152 - 28638	6.9

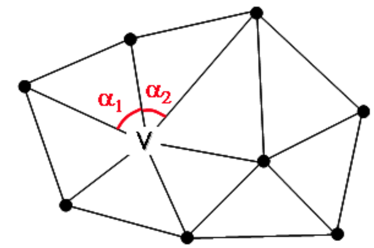
Curvature -- Deviation from the Flat



Curvature

Basic Geometry: Vertex Curvature of Polyhedra

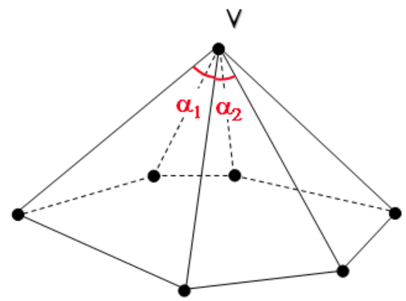
- In the plane, sum of the *face* angles at each (internal) vertex is 2π



- A vertex has “angle defect” when

$$k(v) = 2\pi - \sum_{v \in \, face \, f} \alpha_f > 0$$

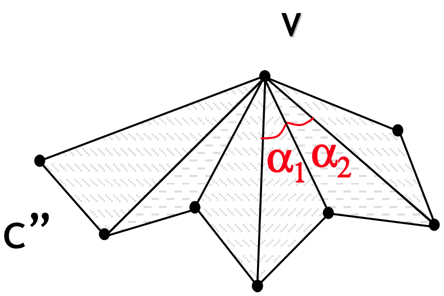
-- positive curvature or “spherica



- A vertex has “angle excess” when

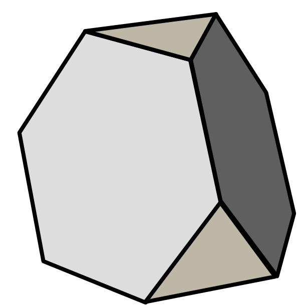
$$k(v) = 2\pi - \sum_{v \in \, face \, f} \alpha_f < 0$$

--negative curvature or “hyperbolic”



- By Descartes’ theorem for polyhedra

$$\sum_{v \in P} k(v) = 2\pi\chi(P)$$

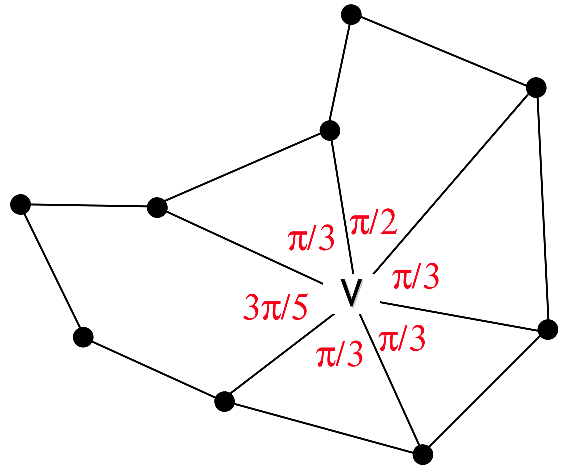


where $\chi(P)$ is the *Euler characteristic* of the polyhedron (=$V - E + F$ =2 if there are no holes and else =2-2g where g is the number of holes)

Combinatorial Vertex Curvature for Planar Graphs

One could imitate the previous definition to define a *combinatorial angular defect/excess* at vertices of a planar graph (net of 2π). E.g.,

$$k(v) = 2\pi - \left(\frac{1}{2} * \frac{2}{4} 2\pi + \frac{\pi}{3} + \frac{1}{2} * \frac{3}{5} 2\pi + \frac{\pi}{3} + \frac{\pi}{3} + \frac{\pi}{3}\right)$$



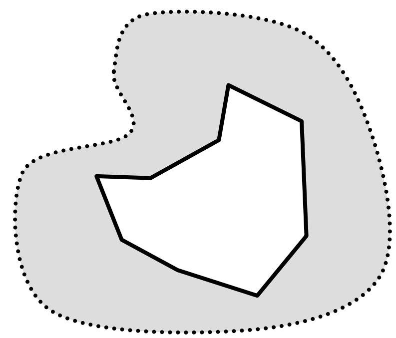
$\chi(G)=2$ if the exterior is counted
$=1$ otherwise

In effect, assume each face is a regular *n-gon*, compute the facial angles, add up and subtract from 2π **[Higuchi'01]**

$$k(v) = 2\pi - \sum_{v \in f} \frac{1}{2} \frac{2\pi(f-2)}{f} = 2\pi\left(1 - \frac{d(v)}{2} + \sum_{v \in f} \frac{1}{f}\right)$$

Gauss-Bonnet theorem (extension of Descartes') then states

$$\sum_{v \in G} k(v) = 2\pi\chi(G) = 2\pi(2-0)$$



The "exterior" face

I. Saniee

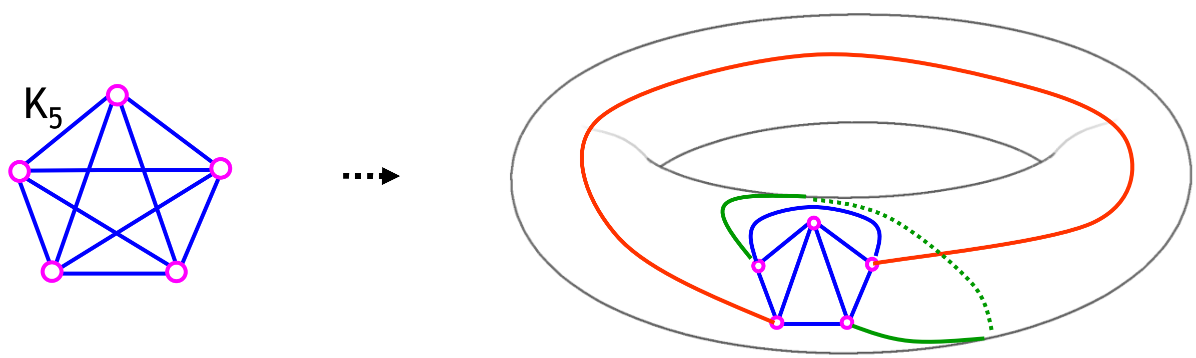
Non-Planar Graphs Always Minimally Embed on Surfaces

What can be said about non-planar graphs? Use the fact that all finite graphs are *locally* planar.

[Ringel-Youngs '68] ("All graphs with N≥3 nodes are locally 2-dimensional.") For N≥3, any $G=(N,L)$ can be embedded in T^g, a torus with g holes, where

$$g \leq \lceil (N-3)(N-4)/12 \rceil$$

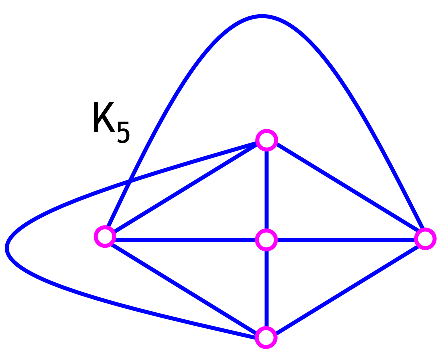
The minimal g is called the *genus* of the graph G.

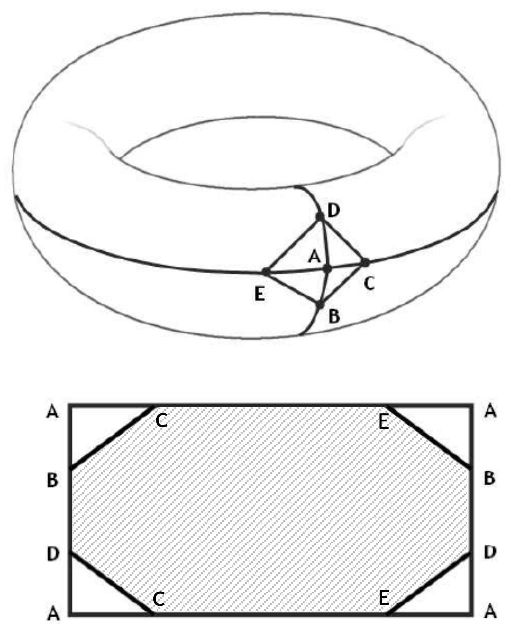


Curvature
Non-Planar Graphs -- Strong Embeddings

But there is more that we need:

[Edmonds-Heffter? see Mohar-Thomassen and others**].** The above embedding can always be done "strongly", i.e., where the resulting embedding on T^g has faces that are 2-cells (equivalent to disks).





B, C, D and E are vertices of the one octagonal face.

Non-Planar Graphs: Combinatorial Curvature

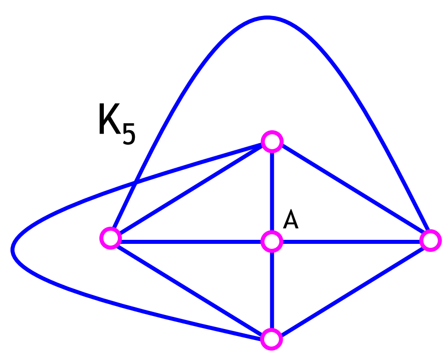
Now with well-defined faces, the previous definition of vertex curvature can be reused:

$$k(v) = 2\pi\left(1 - \frac{d(v)}{2} + \sum_{v \in f} \frac{1}{f}\right)$$

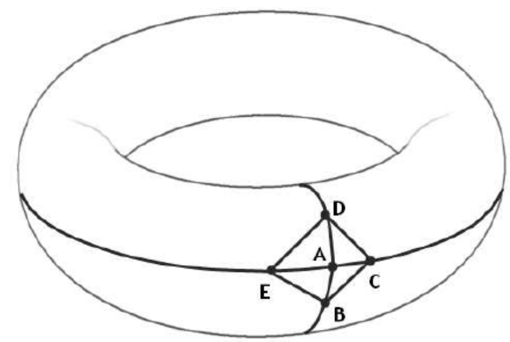
And by Gauss-Bonnet Theorem

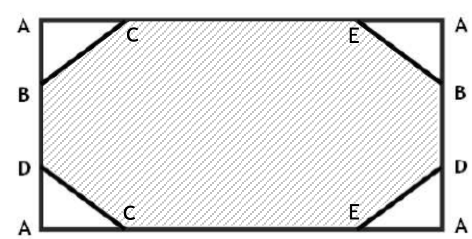
$$\sum_{v \in G} k(v) = 2\pi\chi(G) = 2\pi(2 - 2g)$$

We get the total curvature!



κ(A)=1-4/2+4*1/3=1/3
κ(B,C,D,E)=-1/12





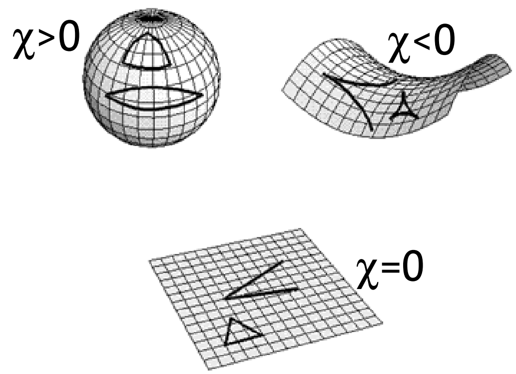
B, C, D and E are vertices of the one octagonal face.

Summary (So Far)

[Exercise] What is the genus of K_7? Identify all faces of the strong minimal embedding of K_7 on T^1. Compute the curvature at each vertex. Verify that $\chi(K_7)$=2-2g.

The **Euler Characteristic** of a graph is an intrinsic invariant that determines its total (combinatorial) curvature*. We say a graph is

- "flat" when $\chi(G)=0$
- "spherical" when $\chi(G)>0$
- "hyperbolic" when $\chi(G)<0$



Note. It is not easy to compute $\chi(G)$ for large scale networks!

* There is also a similar concept of "discrete curvature" for graphs that uses actual edge lengths and angles. It results in the same $\chi(G)$.

Dimension and Curvature

So far:

Managed to define (relatively) satisfactory notions of dimension and curvature for networks but

- **dimension** does not appear to be finite
- **curvature** does not appear to be computable

Give up?

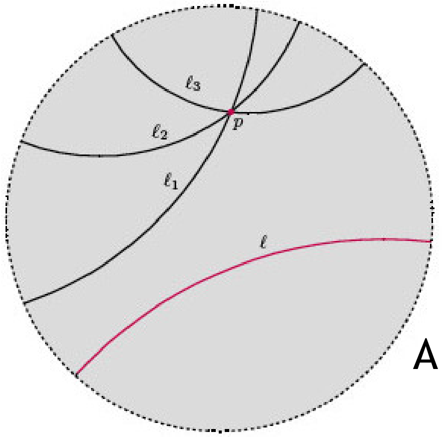
Possible alternative: Consider metric structure of networks

Other Locally 2-Dimensional Models: The Poincaré Disk H^2

Consider the unit disk $\{x \in R^2; |x| < 1\}$ with length metric given by

$$ds^2 = \frac{dx^2 + dy^2}{(1 - x^2 - y^2)^2}$$

the *hyperbolic* metric.



A few geodesics

Advantages

- In *the small scale* it is 2-dimensional, but has much *slower* scaling of geodesics (shortest paths) than √N

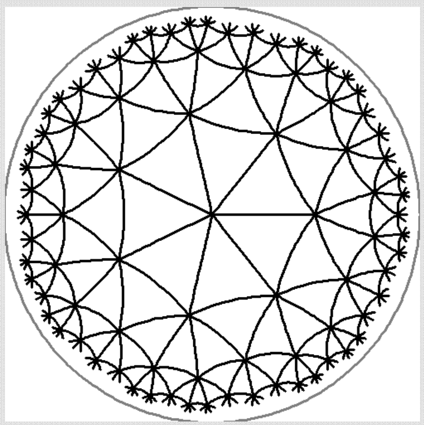
- Has meaningful small-scale and large-scale curvatures

Relationship to graphs? The Poincaré disk comes with numerous natural "scaffoldings" or "tilings".

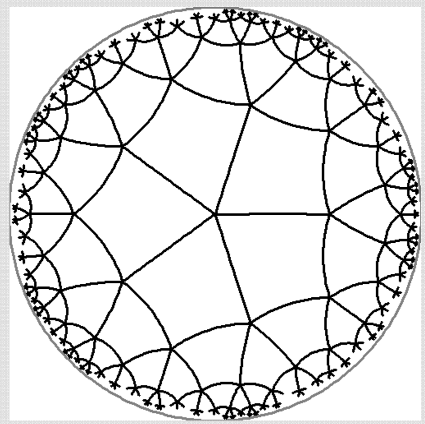
Scaffoldings of H^2: Hyperbolic Regular Graphs

Consider $X_{p,q}$ tilings (isometries) of H^2, that at each vertex consist of *q regular p-gons* for integers p & q with $(p-2)(q-2)>4$ (flat with equality)

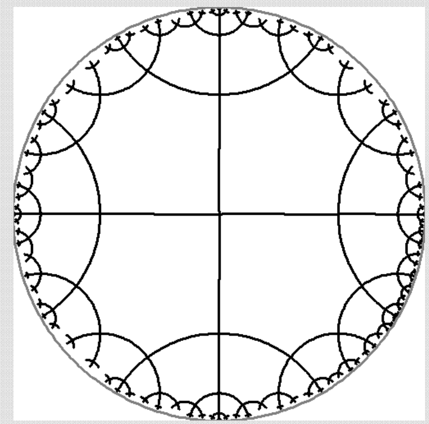
Examples:



$X_{3,7}$



$X_{4,5}$



$X_{6,4}$

Note. Since networks of interest to us are typically finite, we'll consider truncations of $X_{p,q}$, the part within a (large enough) radius r from the center. Call this $TX_{p,q}$.

Some Key Properties of $X_{p,q}$

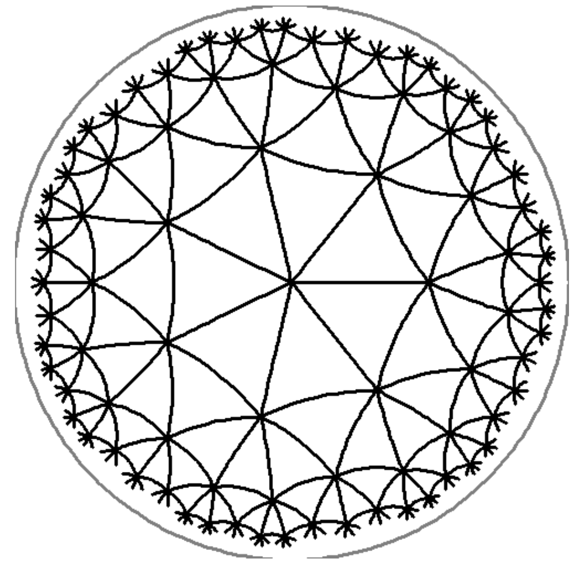
1. **Negative local curvature.** The local combinatorial curvature at each node of $X_{p,q}$ is negative

$$\kappa_v = 2\pi\{1 - \frac{q}{2} + \frac{q}{p}\} = 2\pi\{\frac{4-(p-2)(q-2)}{2p}\} < 0$$

$$\text{or } \kappa_v = 2\pi q(\frac{1}{p} + \frac{1}{q} - \frac{1}{2}) < 0$$

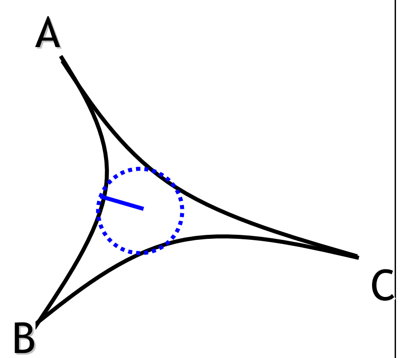
2 . **Exponential growth.** Number of nodes within a ball of radius ***r*** is proportional to λ^r for some $\lambda \equiv \lambda$(p,q) > 1 (e.g., for $X_{3,7}$, $\lambda = \phi$, the golden ratio) or equivalently

2'. **Logarithmic scaling of geodesics.** For (a finite truncation of) $X_{p,q}$ *with N nodes*, the average geodesic (shortest path length) scales like O(log(N))



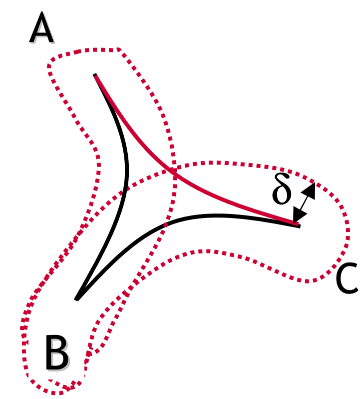
$X_{3,7}$

Curvature in the Large: Geodesic Metric Spaces

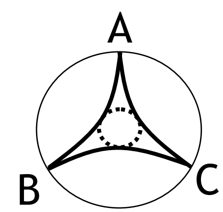


- Computation of *total* curvature of non-flat networks with varying nodal degrees via $\sum_{v\in G}\kappa_v$ does not appear to be possible/easy nor does it provide information about *the large-scale* properties of networks

- A more direct definition of ***(negative) curvature in the large*** is the thin-triangle condition for a geodesic metric space (or a *CAT*(-κ) space):



[M. Gromov's Thin Triangle Condition for a hyperbolic geodesic metric space] There is a (minimal) value $\delta \geq 0$ such that for *any* three nodes of the graph connected to each other by geodesics, each geodesic is within the δ-neighborhood of the union of the other two.



Example. For H^2, $\delta = \ln(\sqrt{2} + 1)$. [Sketch. Largest inscribed circle must be in largest area triangle, $\mathrm{Area}_H(ABC) = \pi - (\alpha+\beta+\chi)$, maximized to π when $\alpha, \beta, \chi = 0$ or when A, B, & C are on the boundary.]

What Can We Say About Communication Networks?

Communication networks are (geodesic) metric spaces via reasonable link metrics (e.g., the hop metric)

Is there evidence for negative curvature in *real* networks?

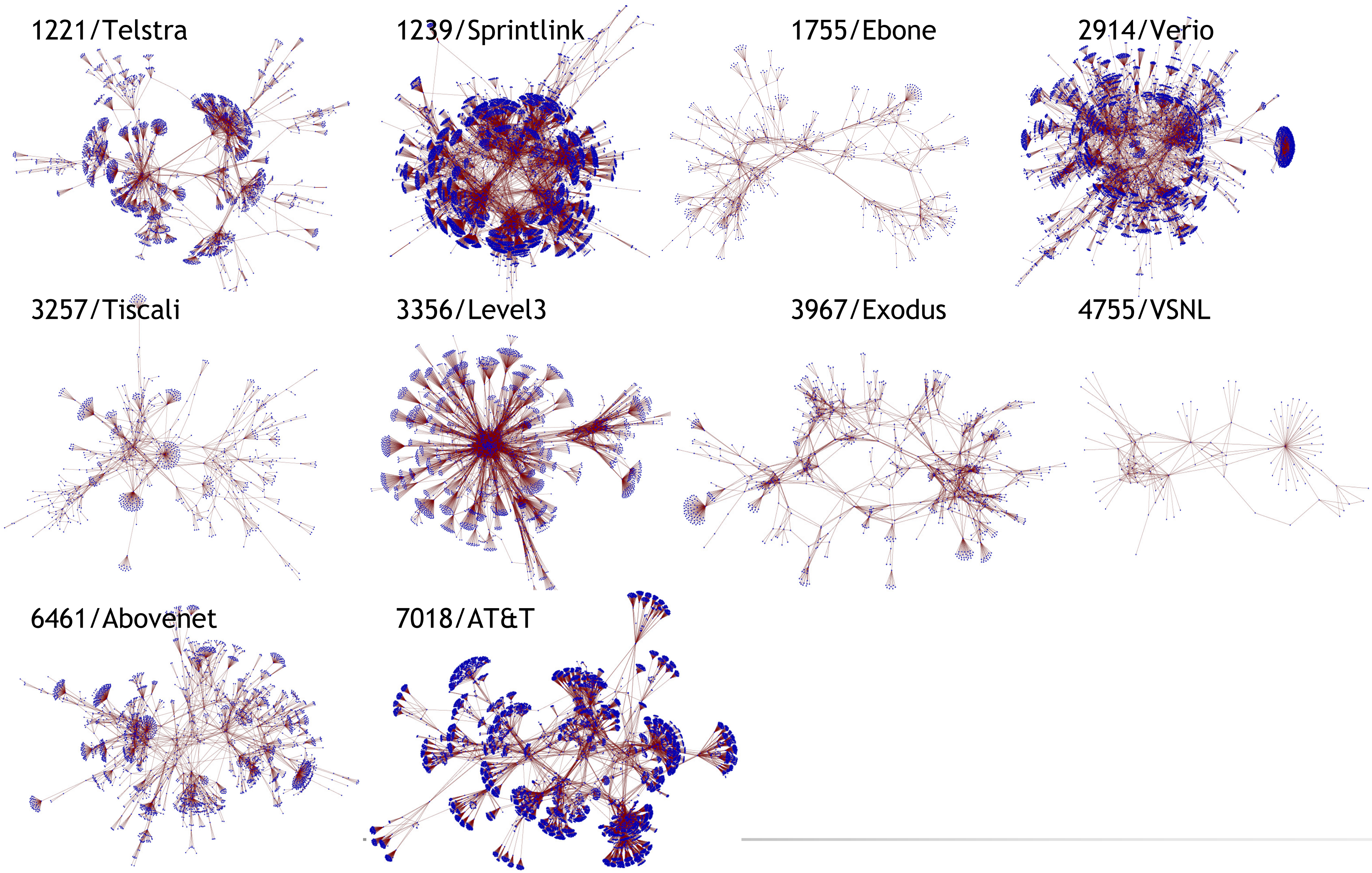
We consider 10 Rocketfuel networks and some prototypically flat or curved *famous* synthetic networks to test this hypothesis

Extracted topologies from RF of 10 global IP networks

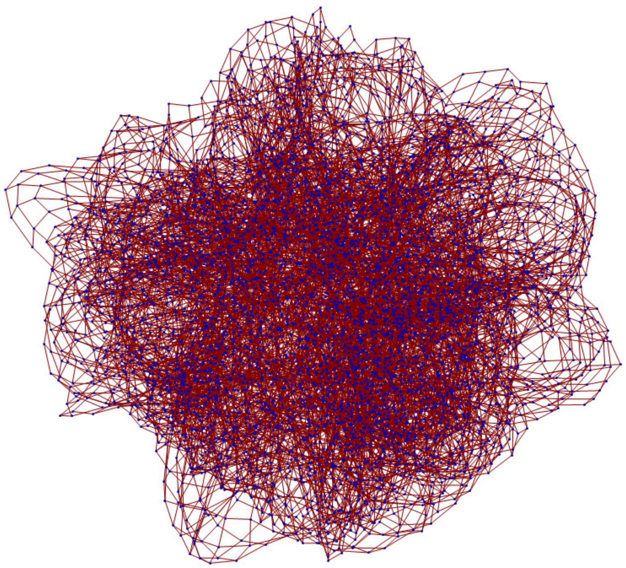
Network ID	Network Name	Number of nodes	Number of links	Diameter
1221	Telstra (Aust.)	2998	3806	12
1239	Sprintlink (US)	8341	14025	13
1755	EBONE (US)	605	1035	13
2914	Verio (US)	3045	12291	13
3257	Tiscali (EU)	855	1173	14
3356	Level 3 (US)	3447	9390	11
3967	Exodus (US)	895	2070	13
4755	VSNL (India)	121	228	6
6461	Abovenet (US)	2720	3824	12
7018	AT&T (US)	10152	14319	12
Hyperbolic 3-7 grid	$X_{3,7}$, synthetic	4264	7511	14
Barabasi-Albert	(B-A), synthetic	10000	19997	9
Watts-Strogatz	(W-S), synthetic(p=0.2)	80x80	13289	20
Triangular lattice	synthetic	469	1260	24
Square lattice	synthetic	80x80	12640	158
Erdos-Renyi	(E-R), synthetic	7992	20132	30

In RF data, a **node** is a unique IP address and a **link** is a (logical) connection between a pair of IP addresses enabled by routers, physical wires between ports, MPLS, etc.

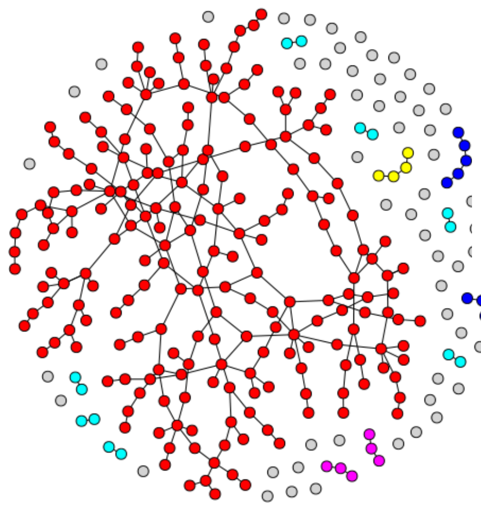
Rocketfuel IP Networks



Some "Famous" Synthetic Networks: E-R, W-S, B-A

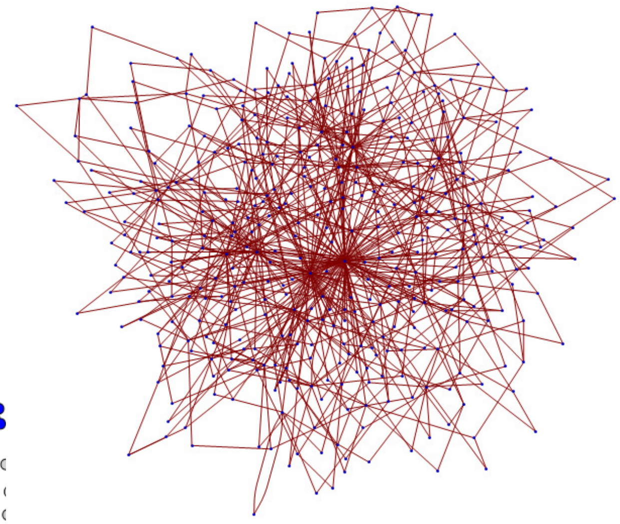


W-S: Grid with small (1-5%) additional random links



G(N,p) random graph

p~1/N, GC~ O(ln(N))
p~ln(N)/N, GC~O(N)



B-A: Start with some nodes and add nodes sequentially and at each iteration join new node to existing node i with probability
$p= d(i)/\sum d(i)$
Then $P(k) \sim k^{-3}$

Experiments and Methodology

We ran experiments on all Rocketfuel networks plus a few prototypical flat/curved networks to test our key hypothesis:

1. Dimension. “Growth test” – Polynomial or exponential?

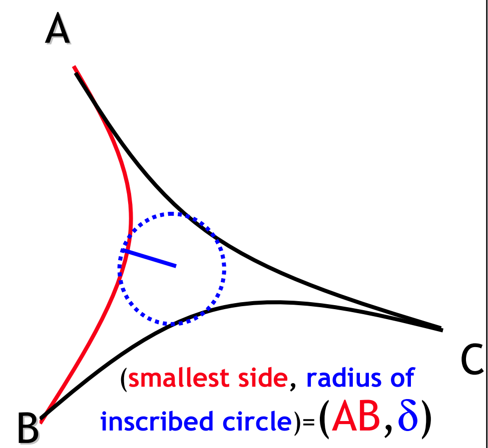
- Consider the volume V(r) as a function of radius r for arbitrary centers

[In flat graphs volume growth is typically polynomial in radius *r*]

2. Curvature. “Triangle test” – Are triangles are universally δ-thin

- Randomly selected 32M, 16M, 1.6M triangles for networks with more than 1K nodes and exhaustively for the remainder
- For each triangle noted shortest side *L* and computed the δ
- Counted number of such triangles, indexed by δ and *L*

[In flat graphs δ grows without bound as the size of the smallest side increases]



We conduct “growth” and “triangle” tests

1. Growth Charts

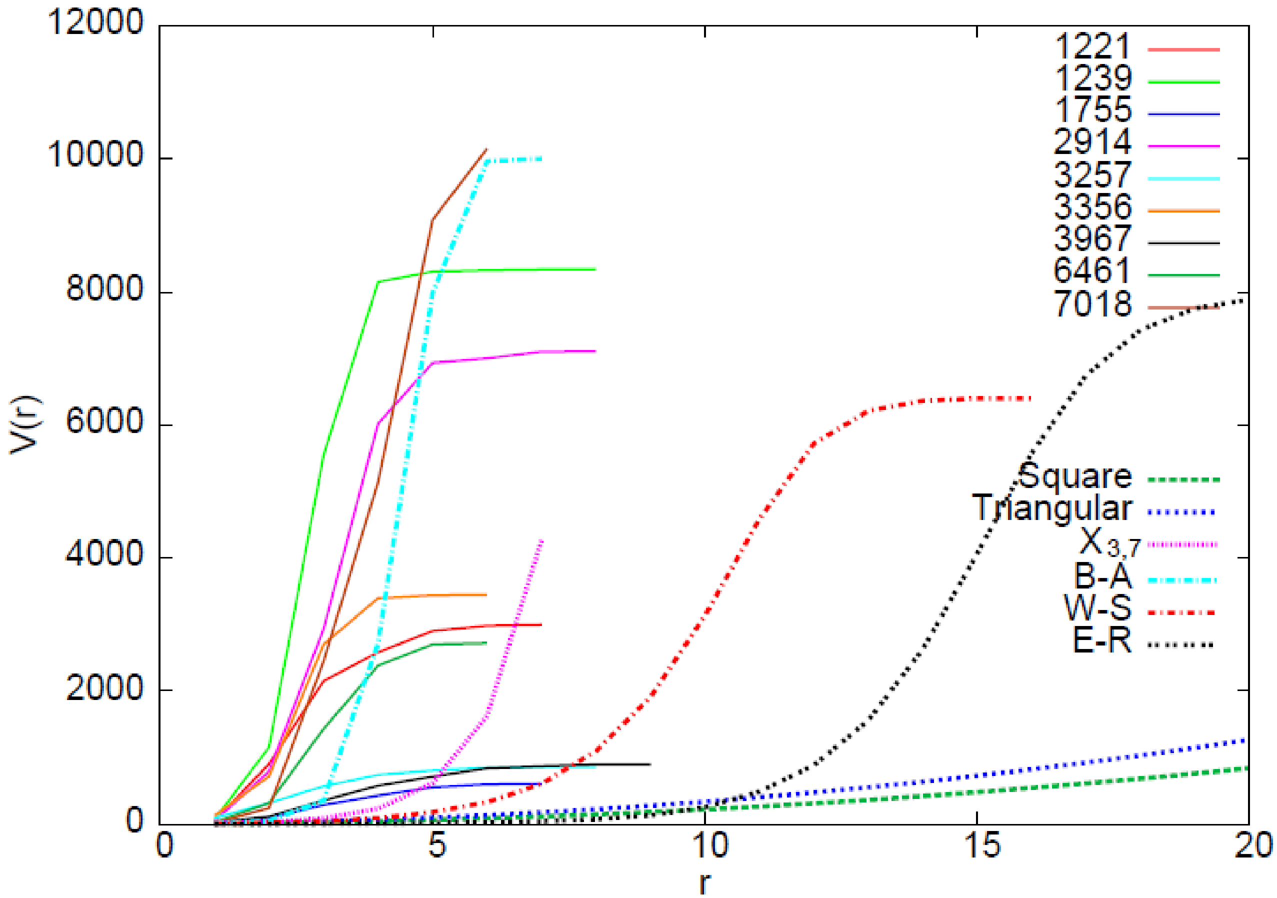
Recall that:

Euclidean growth $V(\boldsymbol{r}) \approx \boldsymbol{r}^D$

then dimension is "D"

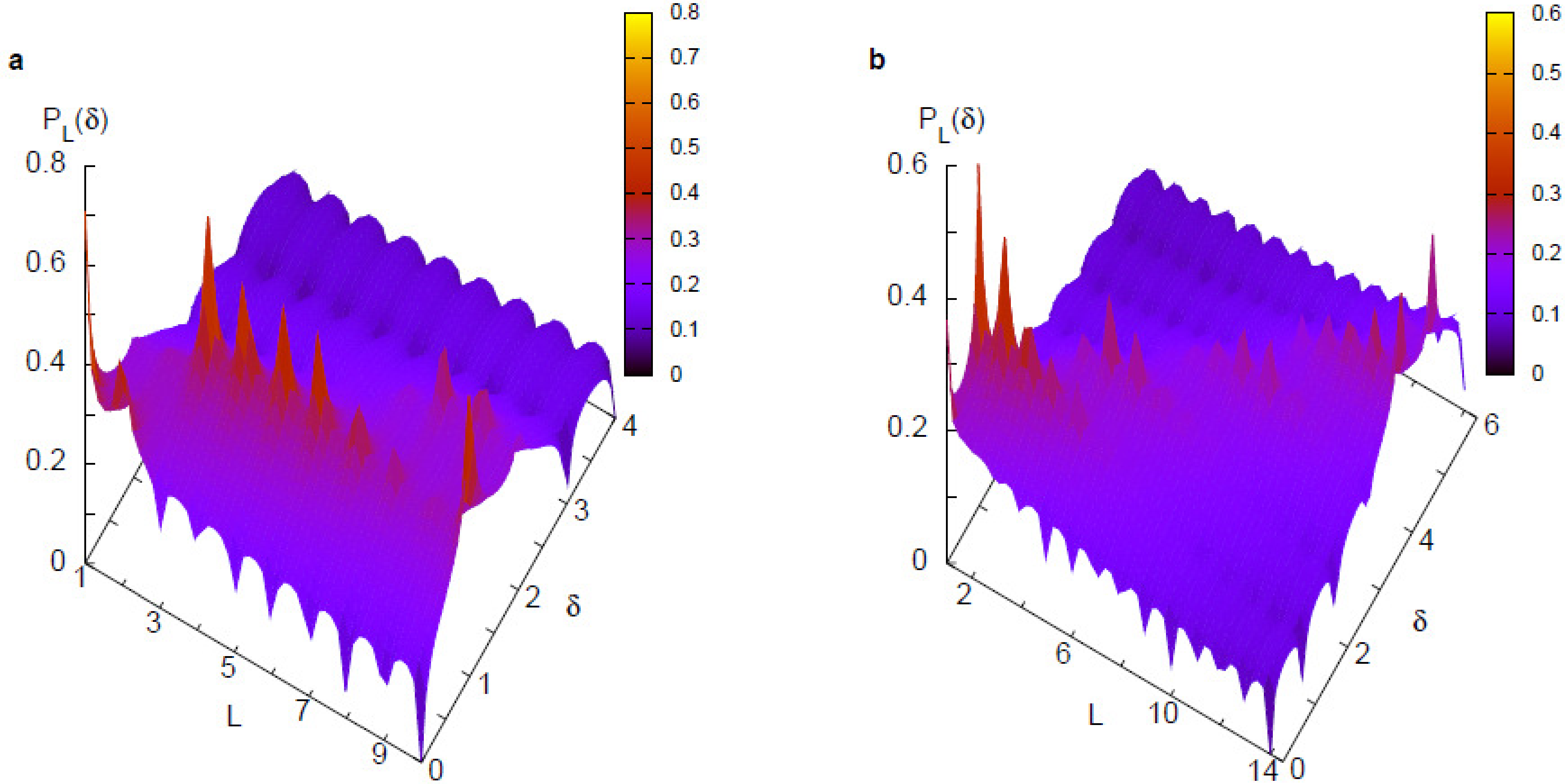
Exponential growth $V(\boldsymbol{r}) \approx \theta^r$

then dimension is "infinity"



Volume (number of points within distance r) as a function of *radius r* from a "center" of the graph. Flattening of curves for larger *r* is due to boundary effects / finite size of network.

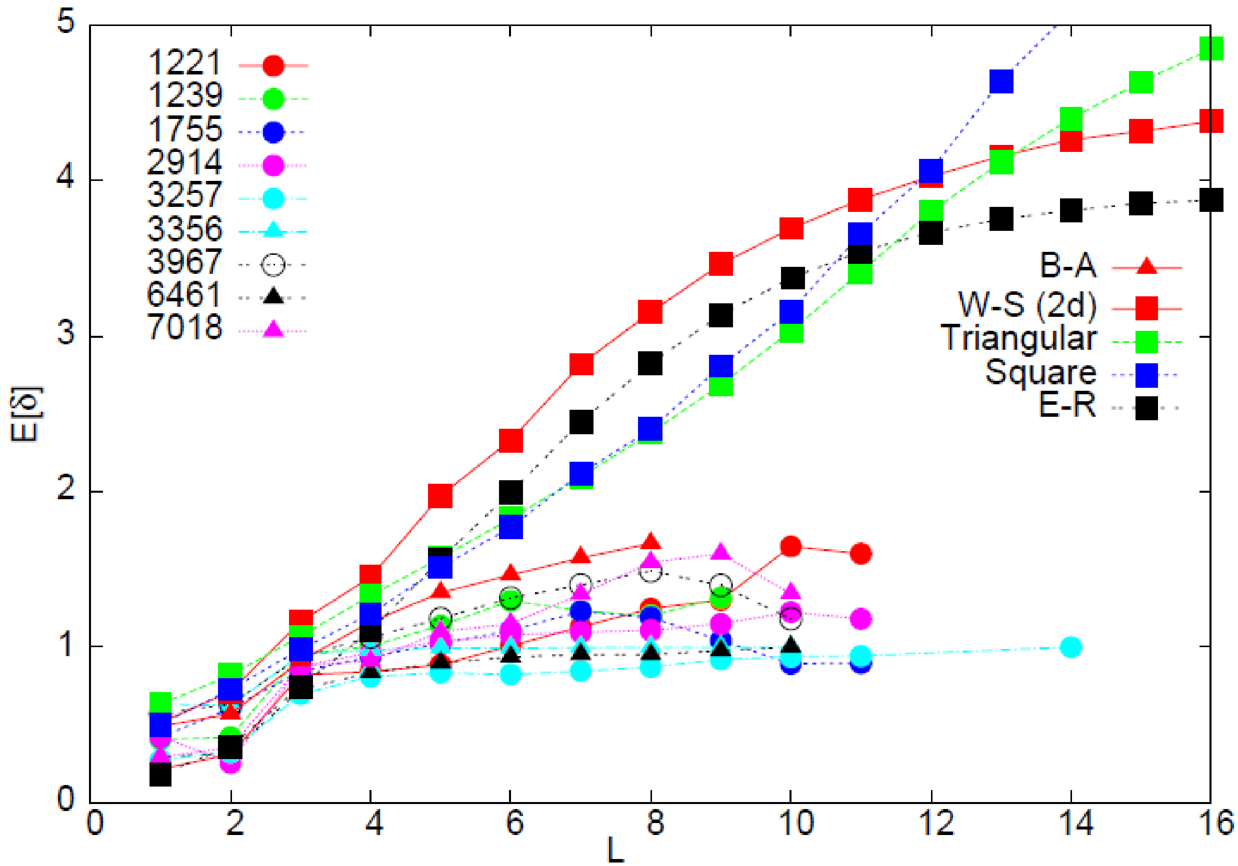
2. Triangle Test - Rocketfuel 7018 & Triangular Grid



(a) Probability $P_L(\delta)$ for randomly chosen triangles whose shortest side is L to have a given δ for the network 7018(AT&T network) which has 10152 nodes and 14319 bi-directional links and diameter 12. The quantities δ and L are restricted to integers, and the smooth plot is by interpolation.

(b) Similar to (a), for a (flat) triangular lattice with 469 nodes and 1260 links. (The smaller number of nodes is sufficient for comparing with (a) since the range for L is large due to the absence of the small world effect.)

Summary of Triangle Tests for Rocketfuel Networks



The average δ as a function of L, E[δ](L), for the 10 IP-layer networks studied here, and for the Barabasi-Albert model with k = 2 and N = 10000 (11th curve) and the hyperbolic grid X3,7 (12th curve). On the other hand, a Watts- Strogatz type model on a square lattice with N = 6400, open boundary conditions and 5% extra random connections (13th curve) and two flat grids (the triangular lattice with diameter 29 and the square lattice with diameter 154) are also shown.

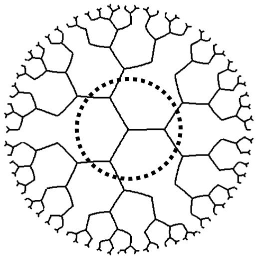
Where to go from here?

- OK, these ten RF datasets and some “well-known” large-scale networks exhibit
  - Exponential growth / logarithmic scaling of shortest paths
  - Negative curvature in the large

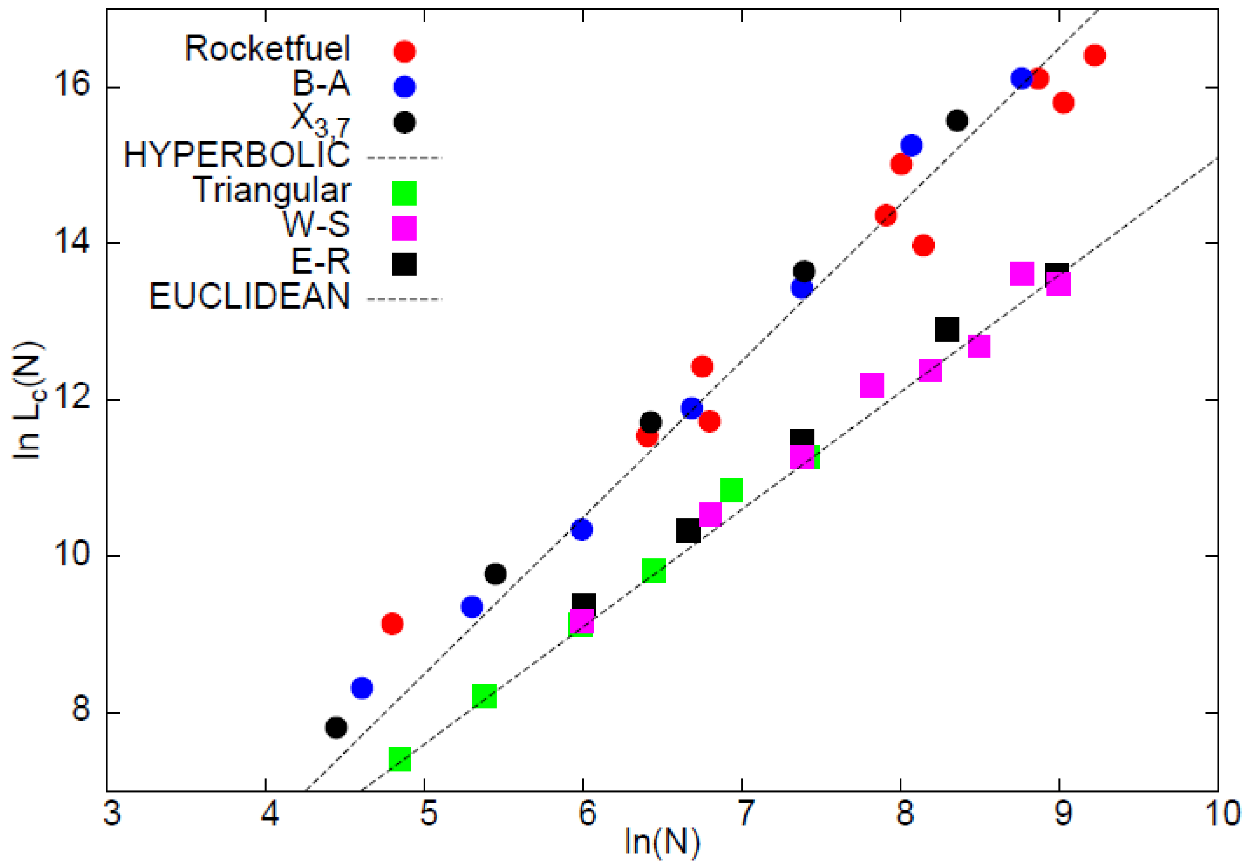
So what?

Turns out negatively-curved networks exhibit specific features that affect their critical properties -- Existence of a “core”:

- $O(N^2)$ scaling of “load” (1 unit between all node pairs)
- Non-random points of critical failures
- Non-random points of security



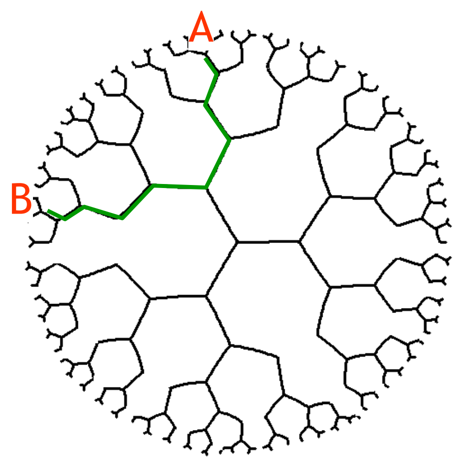
The Downside of Hyperbolicity: Quadratic Scaling of Load ("Betweenness Centrality" and Existence of "Core")



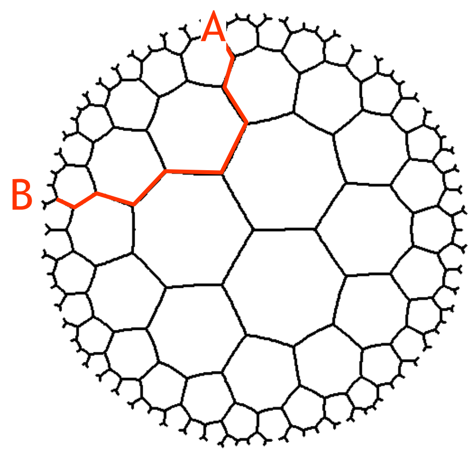
Plot of the maximum load $L_c(N)$ -- maximal number of geodesics intersecting at a node -- for each network in the Rocketfuel database as a function of the number of nodes N in the network. Also shown are the maximum load for the hyperbolic grid X3,7, the Barabasi-Albert model with k = 2, the Watts-Strogatz model and a triangular lattice, for various N. The dashed lines have slopes of 2.0 and 1.5, corresponding to the hyperbolic and Euclidean cases respectively.

Metric Properties of RF and Other networks

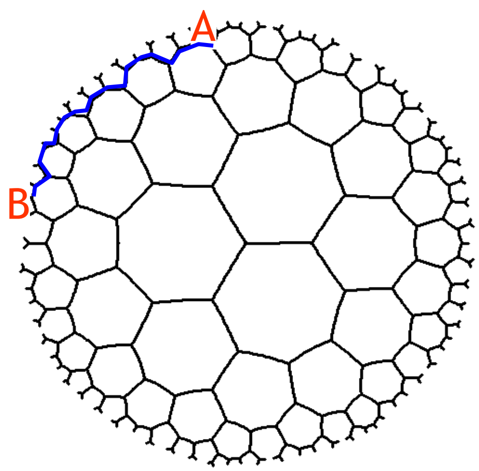
- So far we worked with the unit-cost (hop) metric
- Can things change significantly through changes in the metric?
- Yes, and no! Look at toy networks again:



$T_{3,7}$ with modified hyperbolic metric, load ~ $O(N^2)$



$T_{3,7}$ with hyperbolic metric load ~ $O(N^2)$



$T_{3,7}$ with Euclidean metric load ~ $O(N^2)/\log N$

- Metrics can change things but evidently not by that much! (Need rigorous proofs to determine by how much)

Downside of Metric Changes: Long Paths (w.r.t. the hop metric)

Even if we can eliminate $O(N^2)$ scaling of load via metric changes, we're liable to pay a (big) price:

[Bridson-Haefliger] Let X be a δ-hyperbolic geodesic metric space. Let C be a path in X with end points p and q. Let $[p,q]$ be the geodesic path. Then for every x on $[p,q]$

$$d(x,C) \leq \delta \, \| \log_2 l(C) \| + 1$$

where $l(C)$ is the length of C.

Open Question. Can paths with small deviations from geodesics decrease "load" by much? [Unlikely in the mathematical sense but perhaps yes in practice.]

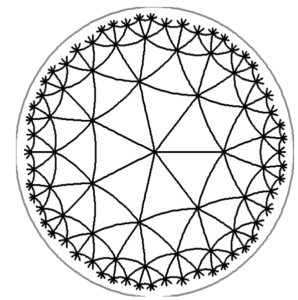
Key Claims:
Network Curvature -> Congestion, Reliability and Security

Numerical studies show that congestion is a property of the large-scale geometry of the networks – large-scale curvature -- and does not necessarily occur at vertices of high degree but rather at the points of high cross-section (the "core")

At the "core" –- intersection of largest number of shortest paths – load scales as quadratic as function of network size

Shortest path routings

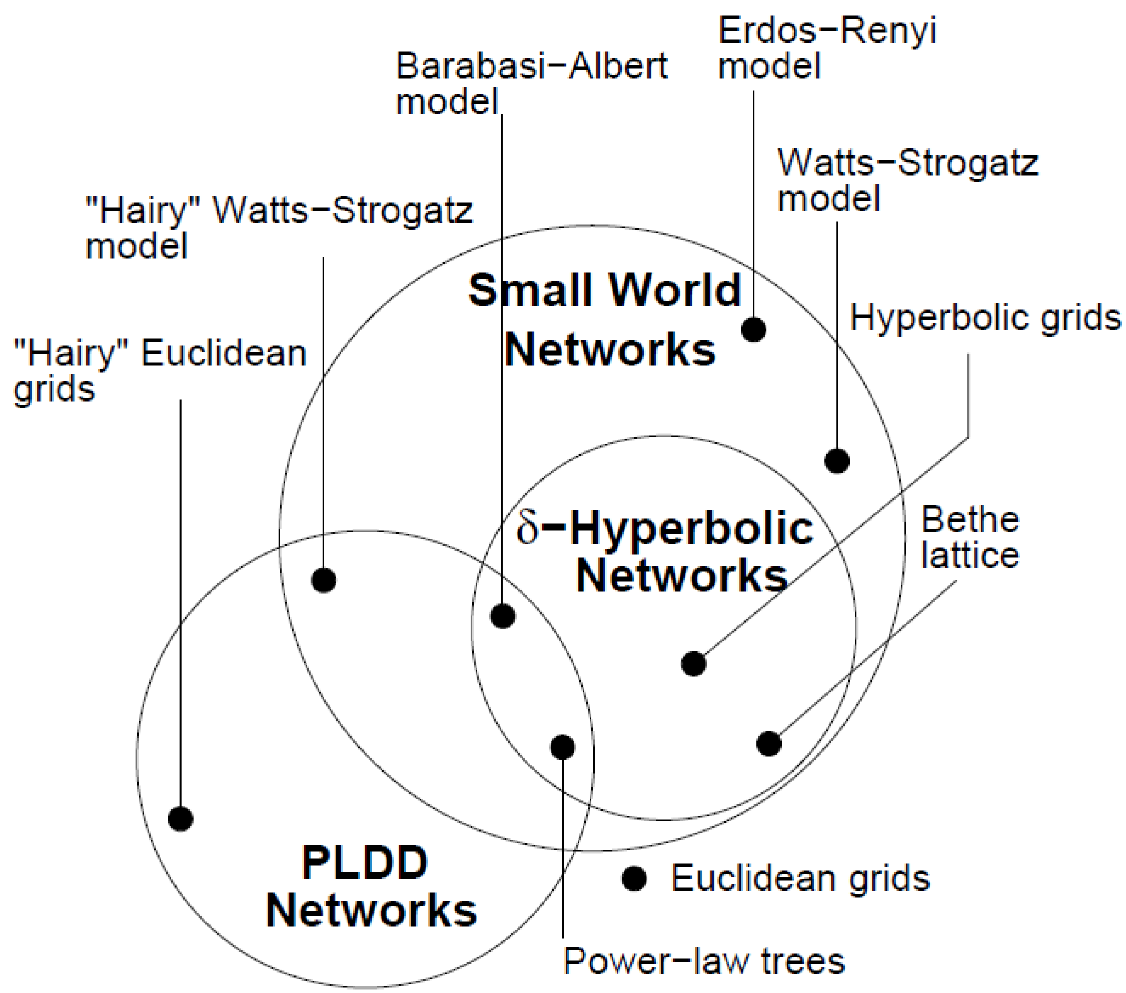
- (Upside) Are very effective, as diameter is small compared to N, e.g., TTL of ~20 good enough for all of the Internet!
- (Downside) Lead to
  - congestion
  - non-random failure can be severe
  - certain nodes exhibit more significant security compromise



$X_{3,7}$

Nodal *loads* need not be related to nodal *degrees*

A Taxonomy for Large-Scale Networks



Taxonomy of key characteristics of networks and their overlaps in a schematic diagram.

CHALLENGES: Impact of Curvature on Metrics Implications for CDNs, Cloud, Virtual Network Design

- Analysis of larger datasets
  - Communication data
  - Bio data
  - Social network data
- Scaling of algorithms for detection of hyperbolicity in much larger graphs (of ~10^9 nodes)
- How does "negative curvature in the large" affect performance, reliability and security?
  - Speed of information/virus spread → spectral properties of large graphs
  - Impact of correlated failures → Core versus non-core
- How does the $O(N^2)$ scaling of load change as a function of alternative load profiles, e.g., for localization in CDNs?
- How $O(N^2)$ affect reliability and security? Does a core add or diminish robustness / security?
- How to leverage hyperbolicity for data centers / cloud / virtualization? Are there fundamental designs?
- How to leverage hyperbolicity for caching and CDNs? DHTs?

Some Recent References

[1] O. Narayan, I. Saniee, *Scaling of Load in Communications Networks*, **Physical Review E**, Sep. 2010.

[2] O. Narayan, I. Saniee, *The Large Scale Curvature of Networks*, **arXiv:0907.1478** (July 2009)

[3] O. Narayan, I. Saniee, G. Tucci, *Lack of spectral gap and hyperbolicity in asymptotic Erdos-Renyi random graphs*, **arXiv:1099.5700v1** (Sep 2010)

[4] Y. Baryshnikov, G. Tucci, *Scaling of Load in delta-Hyperbolic Networks,* **arXiv:1010.3304** (March 2010)

[5] E. Jonckheere, M. Lou, F. Bonahon, Y. Baryshnikov, *Euclidean versus Hyperbolic Congestion in Idealized versus Experimental Networks* **arXiv:0911.2538v1** (Nov. 2009)

[6] Matthew Andrews. *Approximation algorithms for the edge-disjoint paths problem via Raecke decompositions*. FOCS '10