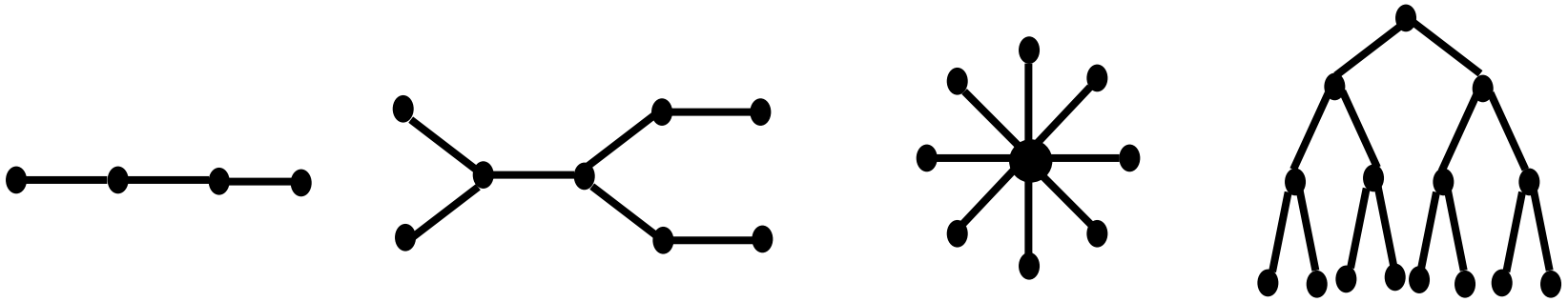
Tree-like structure in social graphs

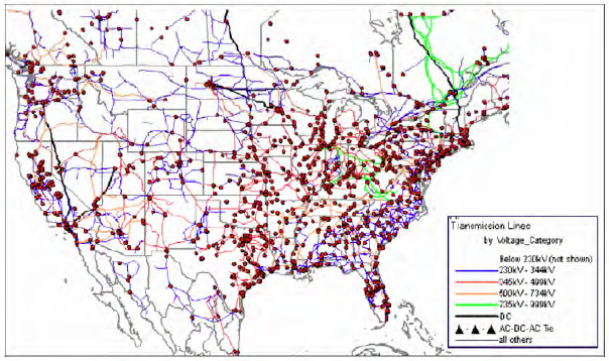
Michael W. Mahoney
ICSI and Dept of Statistics, UC Berkeley

(For more info, see: http://www.stat.berkeley.edu/~mmahoney/)

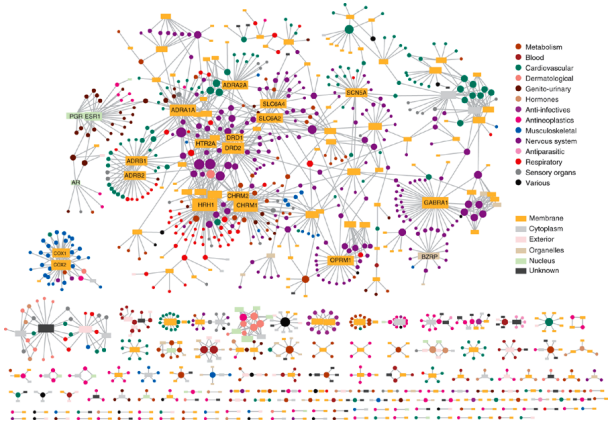


Motivation

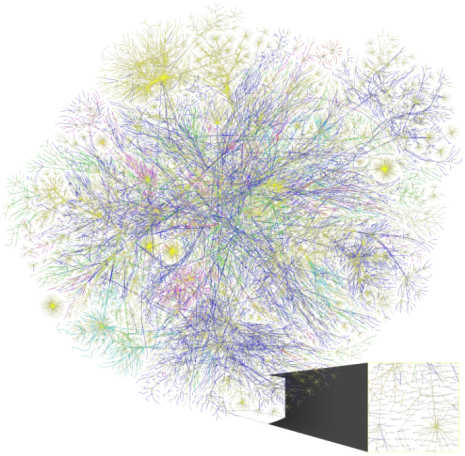
- Graphs/networks: ubiquitous in many domains – e.g. biology, physics, chemistry, infrastructure, communications, and sociology
- Many methods to understand structure at very large-scale (diameter, degree distribution) and very small-scale (clustering coefficient).
- Very few tools to probe intermediate-scale structure (clusters of size 5K in a 5M node graph).
- Can we develop tools to understand and exploit this intermediate-scale structure?



The US electric transmission system.

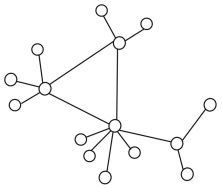


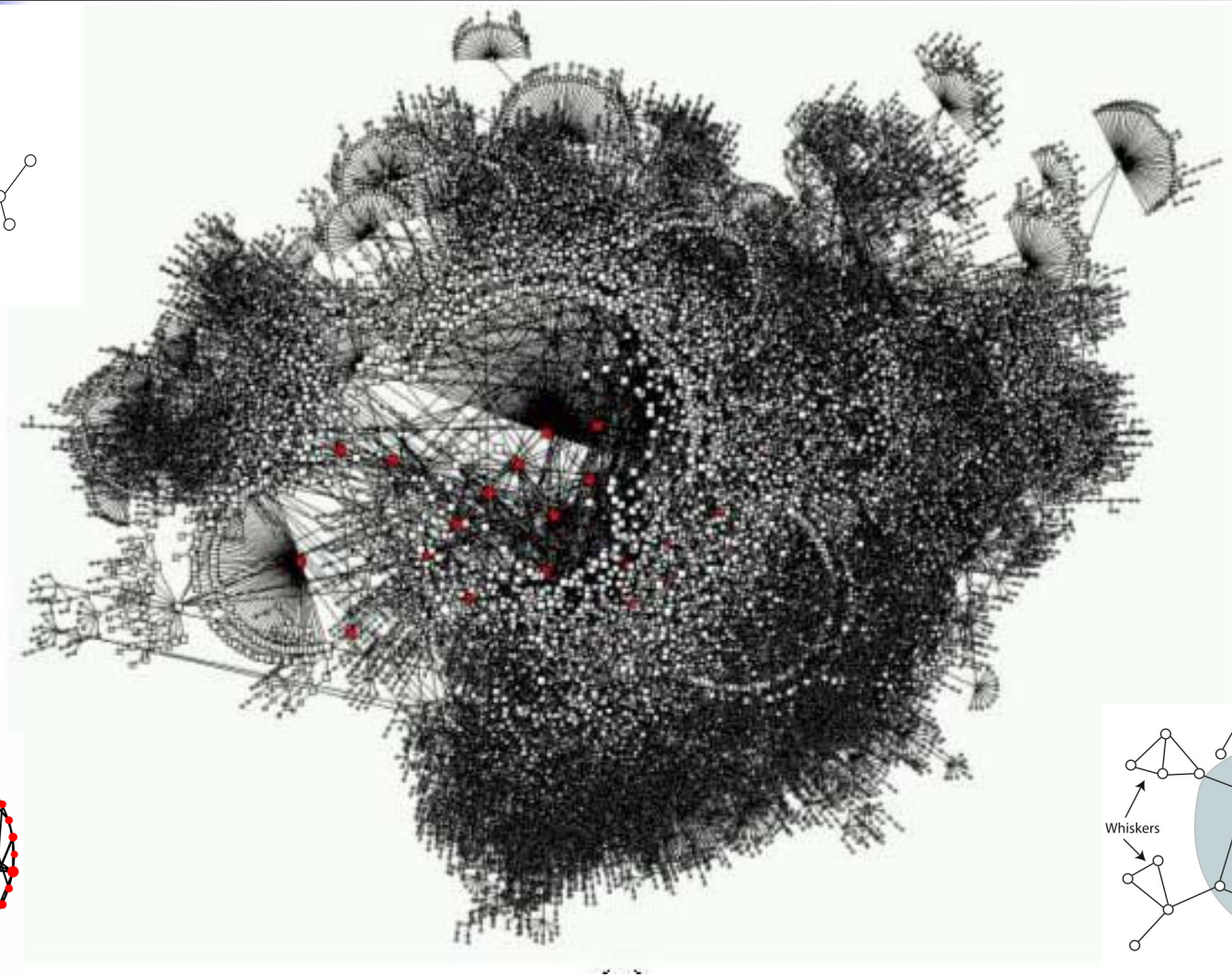
Drug-Target Network.

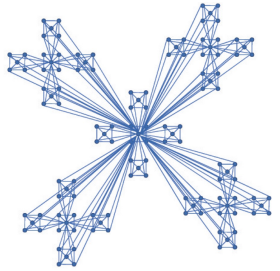


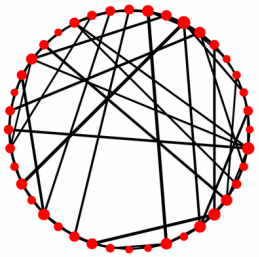
A partial map of the Internet

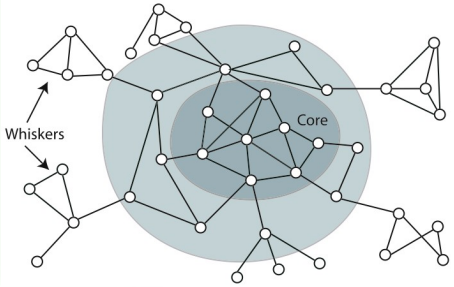
What do social graphs "look like"?

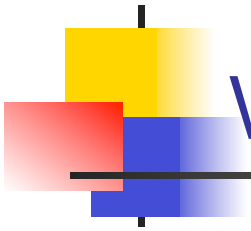












Who cares what social graphs "look like"?

Helpful to develop intuition (for "data scientists")

- Degree distribution, clustering coefficients are very basic stats
- What about better and more robust variants?

Helpful to control inference (for machine learners and statisticians)

- Typically control inference with low-dimensional structures
- What if social graphs are not meaningfully low-dimensional?

Helpful do develop non-trivial theory (for theorists)

- O(lg(n)) distance approximation isn't so good if diameter is O(lg(n))
- Social graphs are very good "hydrogen atom" for development of algorithmic/ statistical methods more generally

But maybe you don't care:

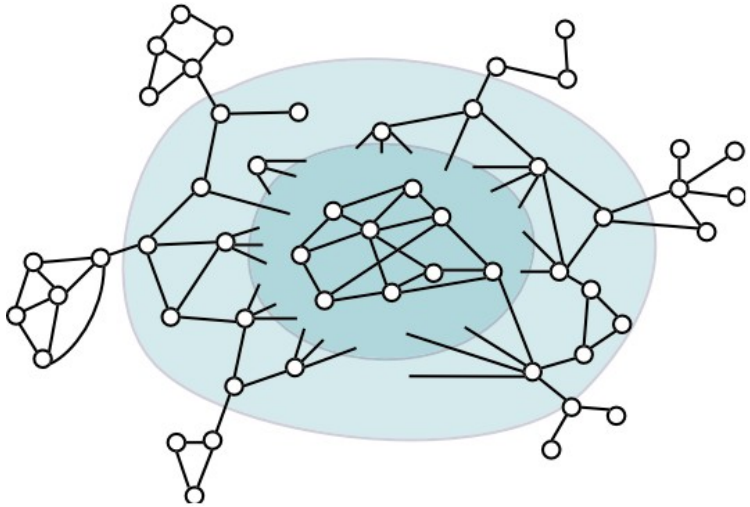
e.g., if you just want to do 0.001% better predicting some well-defined metric.

e.g., if "meta-information" is more important than the graph itself.

How people think about networks

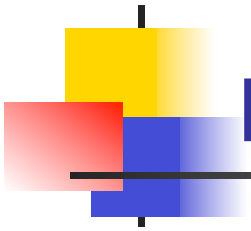
"Interaction graph" ***model*** of networks:

- Nodes represent "entities"
- Edges represent "interaction" between pairs of entities



Graphs are combinatorial, not obviously-geometric

- Strength: powerful framework for analyzing *algorithmic complexity*
- Drawback: geometry used for learning and *statistical inference*



Lots of "networked data" out there!

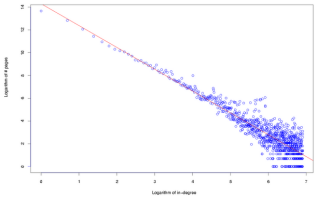
- Technological and communication networks
  - AS, power-grid, road networks
- Biological and genetic networks
  - food-web, protein networks
- Social and information networks
  - collaboration networks, friendships; co-citation, blog cross-postings, advertiser-bidded phrase graphs …
- Financial and economic networks
  - encoding purchase information, financial transactions, etc.
- Language networks
  - semantic networks …
- Data-derived "similarity networks"
  - recently popular in, e.g., "manifold" learning
- …

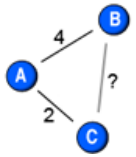
Social and Information Networks

• Social nets	Nodes	Edges	Description
LIVEJOURNAL	4,843,953	42,845,684	Blog friendships [4]
EPINIONS	75,877	405,739	Who-trusts-whom [35]
FLICKR	404,733	2,110,078	Photo sharing [21]
DELICIOUS	147,567	301,921	Collaborative tagging
CA-DBLP	317,080	1,049,866	Co-authorship (CA) [4]
CA-COND-MAT	21,363	91,286	CA cond-mat [25]
• Information networks			
CIT-HEP-TH	27,400	352,021	hep-th citations [13]
BLOG-POSTS	437,305	565,072	Blog post links [28]
• Web graphs			
WEB-GOOGLE	855,802	4,291,352	Web graph Google
WEB-WT10G	1,458,316	6,225,033	TREC WT10G web
• Bipartite affiliation (authors-to-papers) networks			
ATP-DBLP	615,678	944,456	DBLP [25]
ATP-ASTRO-PH	54,498	131,123	Arxiv `astro-ph` [25]
• Internet networks			
AS	6,474	12,572	Autonomous systems
GNUTELLA	62,561	147,878	P2P network [36]

Table 1: Some of the network datasets we studied.

Popular approaches to network analysis



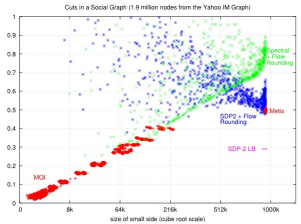


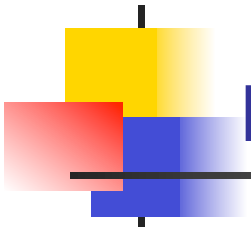
Define simple statistics (clustering coefficient, degree distribution, etc.) and fit simple models

- more complex statistics are too algorithmically complex or statistically rich
- fitting simple stats often doesn't capture what you wanted

Beyond very simple statistics:

- Density, diameter, routing, clustering, communities, …
- Intermediate-scale structure (between very local/small-scale and very global/ large-scale)
- *Popular models often fail egregiously at reproducing more subtle properties (even when fit to simple statistics)*





Failings of "traditional" network approaches

Three examples of *failings of "small world" and "heavy tailed" approaches*:

- Algorithmic decentralized search – (Kle00) - can we *find* short paths?
- Diameter and density versus time – (LKF05) - simple *dynamic* property
- Clustering and community structure – (LLDM08) - subtle/complex *static* property (ubiquitous in downstream analysis) related to *inference*

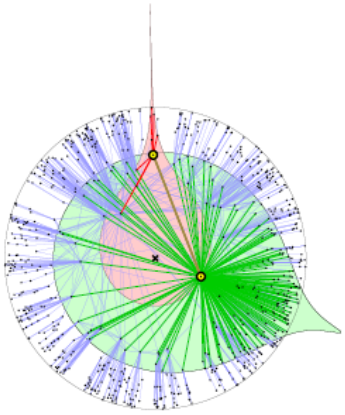
All three examples have to do with the **coupling b/w "local" structure and "global" structure** --- *solution* goes *beyond simple statistics of traditional approaches*.

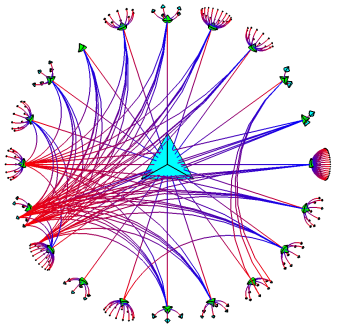
Prior evidence for tree-like structure

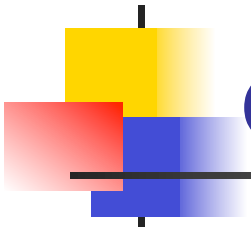
Prior work suggesting graphs/networks may be "tree-like" when viewed at intermediate size scales:

- ***Hyperbolic geometry of complex networks:*** use tree-based models for identifying structure, including a temperature parameter
- ***Geographic Routing Using Hyperbolic Space:*** advantages to embedding internet routing networks into hyperbolic metrics
- ***Community Structure in Large Networks: Natural Cluster Sizes and the Absence of Large Well-Defined Clusters:*** no good large conductance-based clusters, suggests expander-ness with degree heterogeneity enhances hyperbolicity
- ***Complex Graphs and Networks:*** theoretical work on random graphs; the core of "power-law" graphs is a few high-degree, well-connected nodes crucial in many paths
- ***On interactive visualization of high-dimensional data using the hyperbolic plane:*** visualization of large networks in structured hyperbolic metrics like the Poincare disk

However, no consensus has been reached on defining and measuring this tree-like structure, making it difficult to exploit algorithmically.







Outline

Background & failure of simple statistics

- Intermediate-scale structure

Notions of tree-like-ness & general thoughts on core-periphery

- Metric-based versus cut-based versus k-core-based notions of tree-like-ness

δ-hyperbolicity in practice and in theory

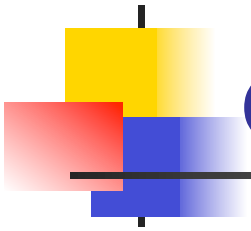
- Many real graphs have size-resolved hyperbolic properties
- Too much randomness in models for theory to be nontrivial

Tree decompositions

- Many practical heuristics developed in scientific computing
- Tool for identifying small-scale and large-scale structure in real graphs

Connections with prior work

- On core-periphery & size-resolved isoperimetric/clustering/community structure



Outline

Background & failure of simple statistics

- Intermediate-scale structure

Notions of tree-like-ness & general thoughts on core-periphery

- Metric-based versus cut-based versus k-core-based notions of tree-like-ness

δ-hyperbolicity in practice and in theory

- Many real graphs have size-resolved hyperbolic properties
- Too much randomness in models for theory to be nontrivial

Tree decompositions

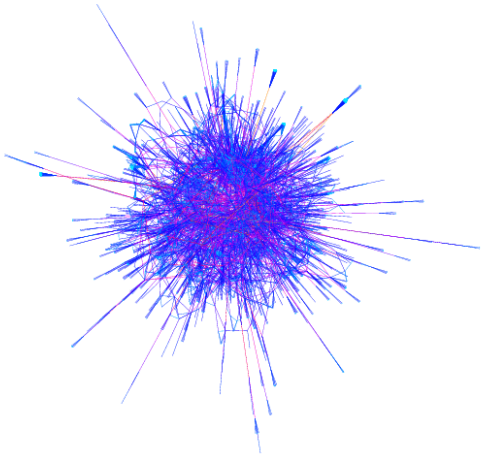
- Many practical heuristics developed in scientific computing
- Tool for identifying small-scale and large-scale structure in real graphs

Connections with prior work

- On core-periphery & size-resolved isoperimetric/clustering/community structure

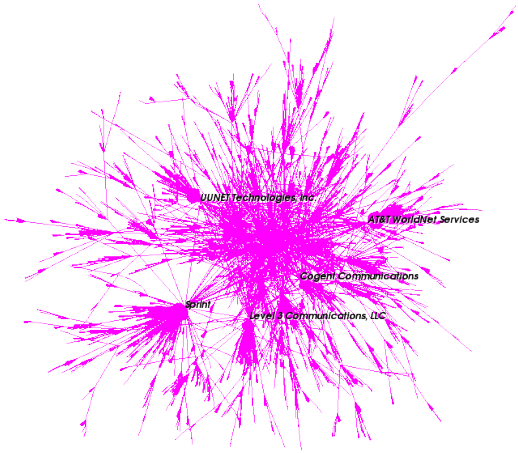
What does "tree-like" mean?

Intuitively: Empirical quasi-randomness (e.g., noise in data or a random graph model) coupled with structural heterogeneities (e.g., variable degree distribution) leads to "tree-like-ness"

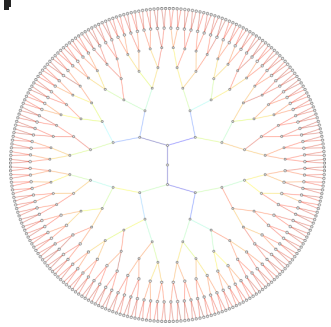


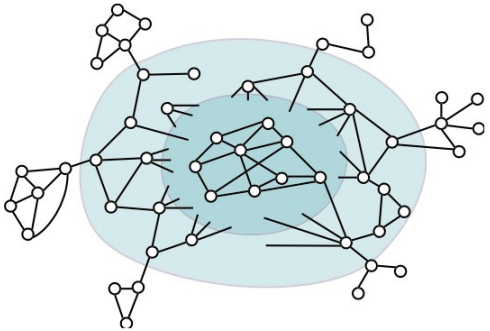
Goal: **Quantify** and **exploit** this idea:

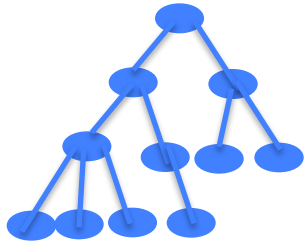
- In a way that is statistically meaningful?
- In a way that is algorithmically tractable?
- In a way that is useful for practitioners in the domain from which the networks are constructed?

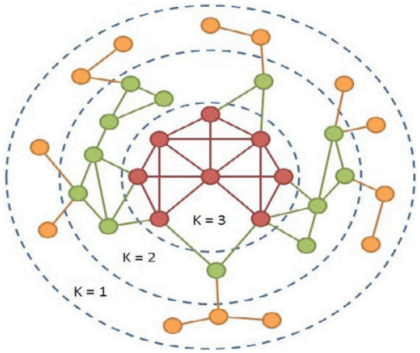


What does "tree-like" mean?

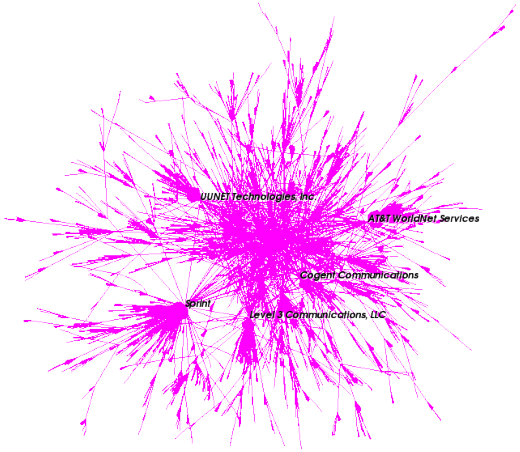








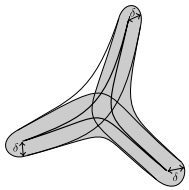
"Although large informatics graphs such as social and information networks are often thought of as having hierarchical or tree-like structure, this assumption is rarely tested, and it has proven difficult to exploit this idea in practice. Moreover, given recent work demonstrating that large informatics graphs have properties that are very different than small social networks and graphs that arise in other machine learning and data analysis applications … it is not clear whether such structure can be exploited for improved graph mining and machine learning, even assuming it exists." [ASM13]



How to measure tree-like-ness

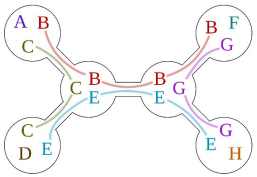
Number of edges to remove to make the graph a tree:

- Not good



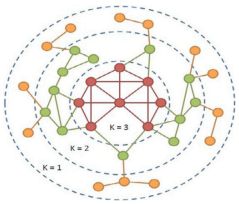
δ-hyperbolicity:

- Notion from geometric group theory that captures tree-like i.t.o. metric structure
- (Existing random graph models (and real data) have "too much randomness" making the use of hyperbolicity very sensitive)



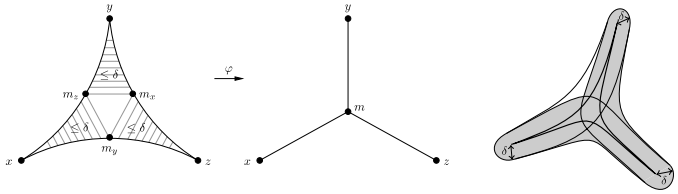
Tree decompositions:

- Notion from structural graph theory that captures tree-like i.t.o. the cut structure
- (Existing heuristics designed for scientific computing often don't perform very well, but they do capture some properties)



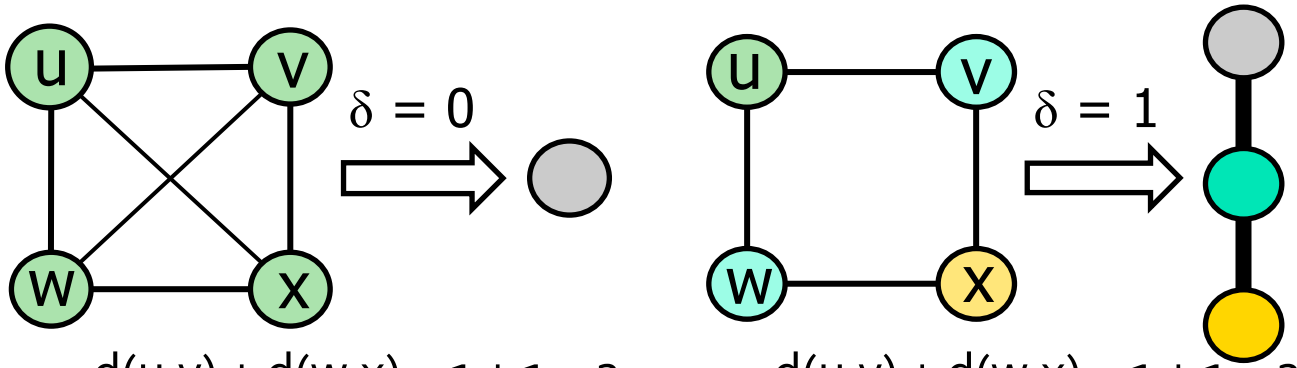
k-core decompositions:

- Fast heuristic that does *not* measure tree-like-ness in general
- (For realistic graphs, it can be "rationalized" in terms of nested-core-periphery structure that tree decompositions and δ-hyperbolicity (different in general) identify)



A geometric measure of tree-like-ness

- Gromov's **δ-hyperbolicity:** arises from metric space geometry; δ measures the extent to which a (geodesic) metric space embeds in a tree metric.



d(u,v) + d(w,x) = 1 + 1 = 2
d(u,x) + d(v,w) = 1 + 1 = 2
d(u,w) + d(v,x) = 1 + 1 = 2

d(u,v) + d(w,x) = 1 + 1 = 2
d(u,x) + d(v,w) = 2 + 2 = 4
d(u,w) + d(v,x) = 1 + 1 = 2

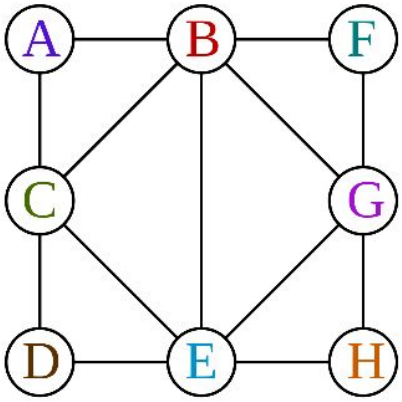
- Note: d(u,v) is the length of shortest path between u and v in G.

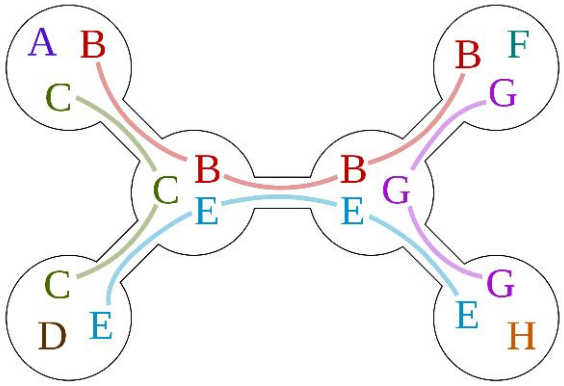
Definition: [Gromov, 1987] A graph is δ-*hyperbolic* iff: For every 4 vertices u, v, w, and z, the larger 2 of the 3 distance sums, $d(u,v)+d(w,z)$ and $d(u,w)+d(v,z)$ and $d(u,z)+d(v,w)$, differ by at most 2δ.

- Several equivalent (up to constants) definitions---δ thin triangles, δ slim triangles, etc.

A combinatorial measure of tree-like-ness

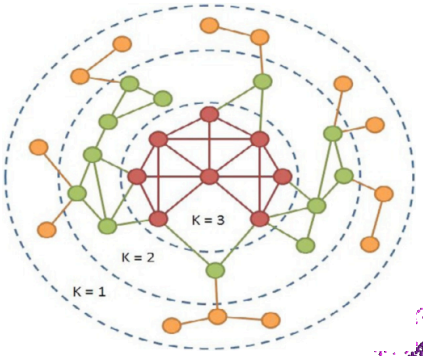
- A **tree decomposition** of a graph $G = (V,E)$ is a pair (X,T), where $X=\{X_1, X_2, \ldots, X_L\}$ is a collection of subsets of V, and T is a tree with nodes $\{1, \ldots, L\}$ satisfying three conditions:
  - The union of the sets in X is equal to V
  - For every edge (u,v) in G, $\{u,v\}$ is a subset of some X_i
  - For every v in V, the indices of $\{X_i\}$ containing V form a sub-tree of T.





- Call the sets X_i the **bags** of the decomposition and $max(|X_i|)$ the **width**. The **tree-width** of **G** is the min width over all valid tree decompositions

Core-periphery: K-core decompositions



- The **k-core** of a graph $G = (V,E)$, denoted $\boldsymbol{H_k}$ is the maximal subgraph H of G s.t. $deg_H(v) \geq k$ for all v in H.
- The **core number** of a vertex v is defined to be the maximum k so that v is in H_k but not H_{k+1}.
- The the **k-shell** of G is the set of nodes with core number k.
- (k-cores do NOT capture tree-like structure in general.)
- (Prior work has shown "nested core-periphery structure" for a different notion of core/periphery.)

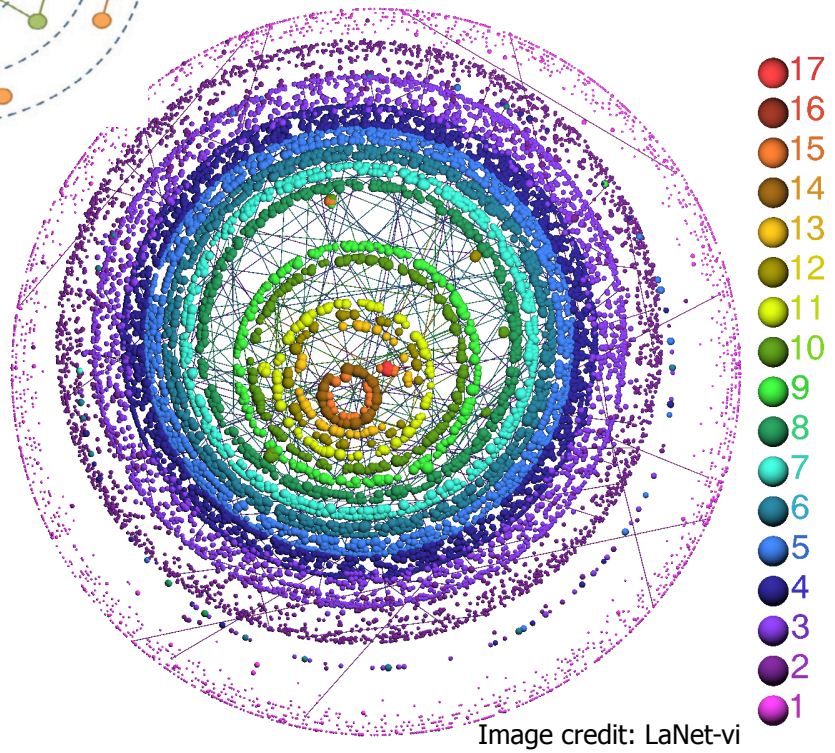


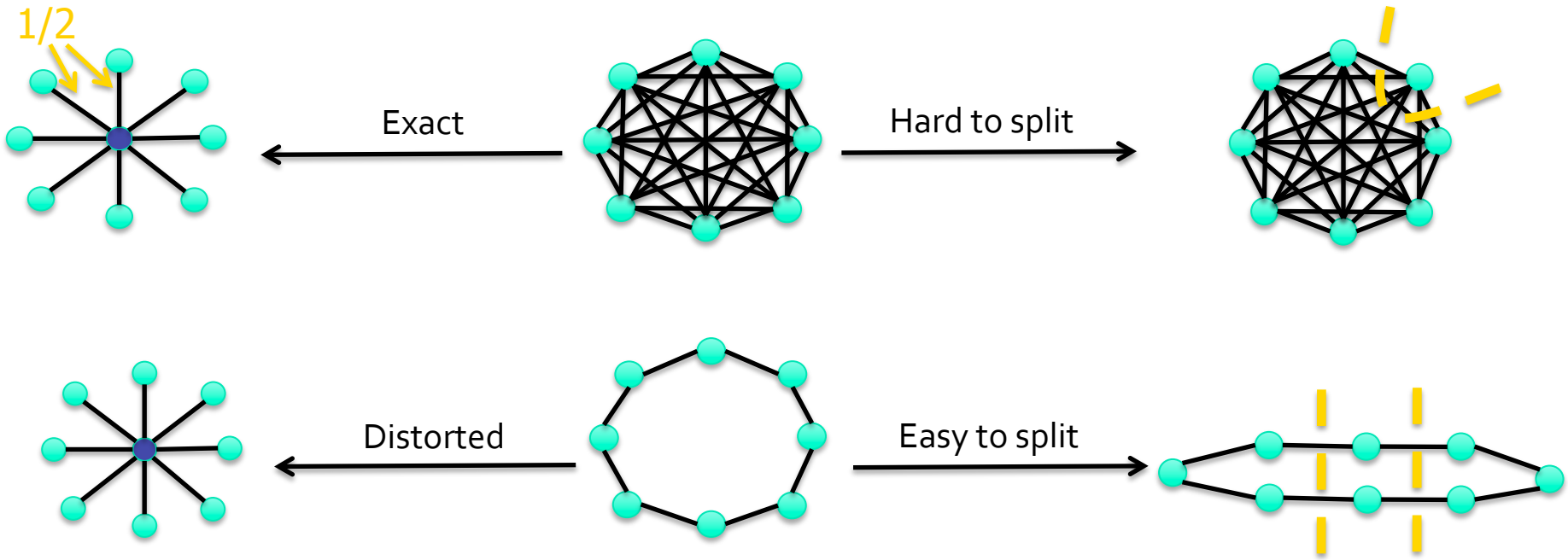
Image credit: LaNet-vi

Condensed Matter Collaboration Network

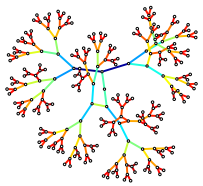
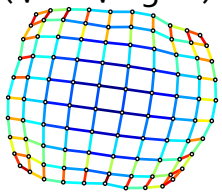
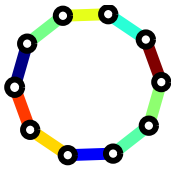
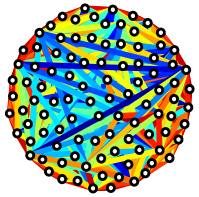
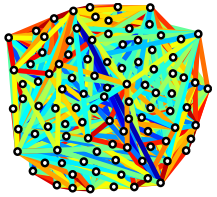
Tree-like-ness: metrics and cuts

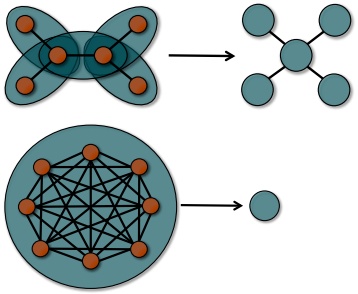
δ hyperbolicity: How similar is the metric structure of the graph to the metric structure of a tree?

Tree decompositions: How similar is the cut structure of the graph to the cut structure of a tree?



Tree-like-ness and toy graphs

	SmallBinary	SmallPlanar (√n x √n grid)	SmallRing	SmallClique	SmallER (p>lg(n))/n)
					
δ (metric-based tree-like-ness)	0	√n-1	n/4	0	lg(n)
Treewidth (but-based tree-like-ness)	1	√n	2	n-1	n



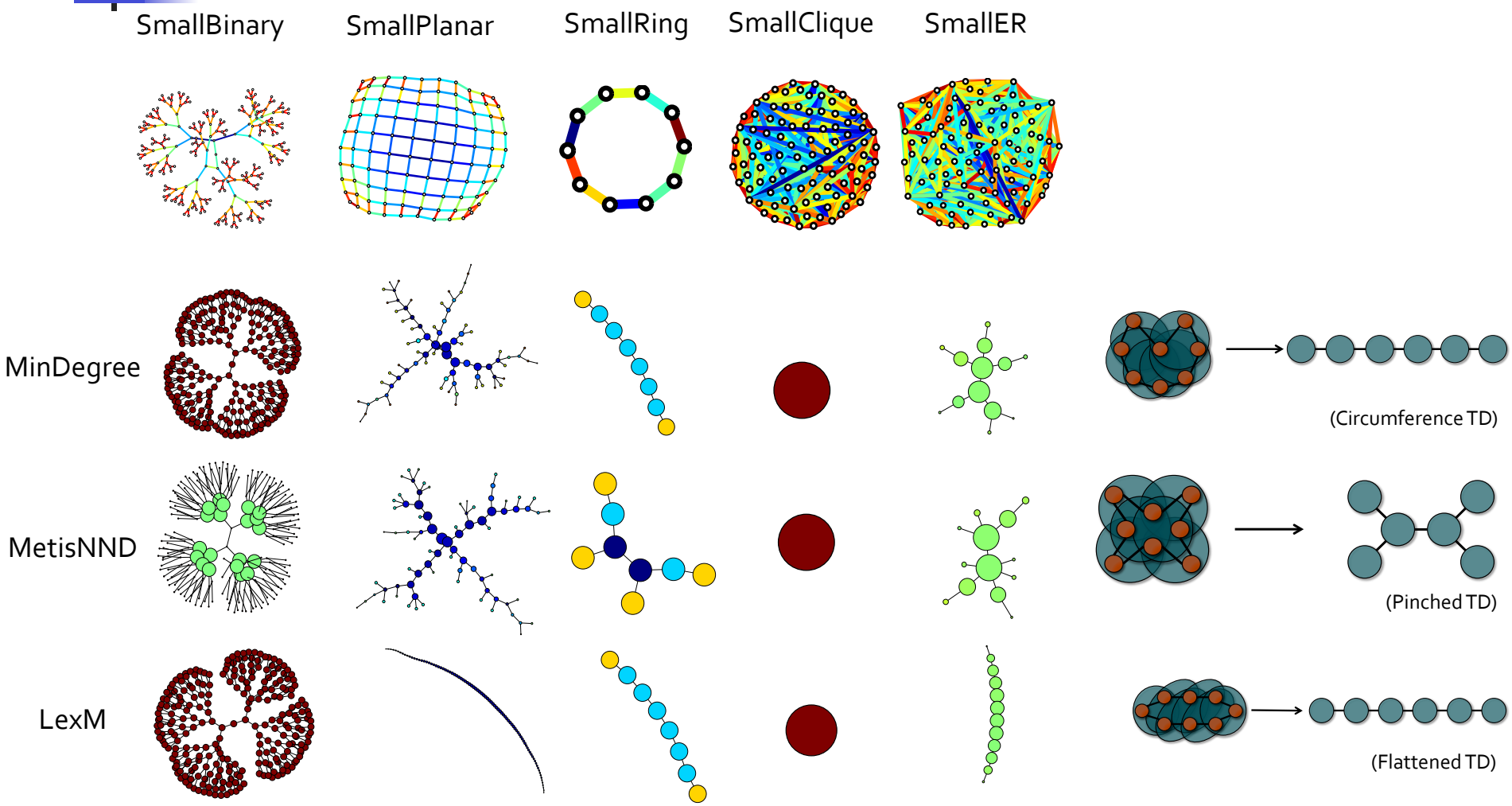
√n, n/4, & lg(n) are all "large": since the equal the diameter

√n, n-1, & n are all "large": since the equal the size of trivial separator

δ and treewidth are maximally different on rings and cliques, but:

- capture similar properties on realistic social graphs
- interesting connections with spectral methods embedding data in lines and cliques

Tree decompositions and toy graphs



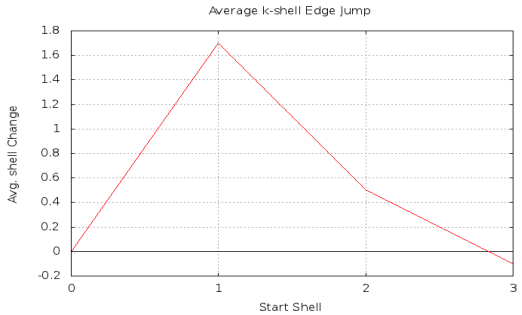
Graphs we look at and why we look at them

Network	n	e	n_c	e_c	$\bar{d}$	$\bar{C}$	D	$\bar{D}$	Description
ER Random Graphs									
ER(1.6)	5000	3996	3210	3471	2.16	0.00	38	15.8	ER graph with $p = 1.6/n$
ER(1.8)	5000	4486	3617	4118	2.28	9.30×10^{-4}	34	12.7	ER graph with $p = 1.8/n$
ER(2)	5000	4986	4001	4783	2.39	9.11×10^{-4}	30	11.9	ER graph with $p = 2/n$
ER(4)	5000	9881	4879	9878	4.05	8.96×10^{-3}	15	6.80	ER graph with $p = 4/n$
ER(8)	5000	20102	4998	20102	8.04	1.59×10^{-3}	7	4.81	ER graph with $p = 8/n$
ER(16)	5000	40215	5000	40215	16.1	3.13×10^{-3}	5	3.86	ER graph with $p = 16/n$
ER(32)	5000	80258	5000	80258	32.1	6.39×10^{-3}	4	3.05	ER graph with $p = 32/n$
PL Random Graphs									
PL(2.25)	5000	5790	3393	5634	3.32	.0131	16	5.51	PL graph with $\gamma = 2.25$
PL(2.50)	5000	7238	4895	6802	2.78	2.46×10^{-3}	18	6.65	PL graph with $\gamma = 2.50$
PL(2.75)	5000	6236	4650	5641	2.43	6.99×10^{-4}	22	8.20	PL graph with $\gamma = 2.75$
PL(3.00)	5000	5363	4071	4556	2.24	1.18×10^{-3}	29	10.1	PL graph with $\gamma = 5.00$
SNAP Graphs									
AS20000102	6474	12572	6474	12572	3.88	.399	9	4.34	Snapshot of autonomous systems network
CA-GrQc	5241	14484	4158	13422	6.46	.665	17	6.74	Collaboration network, general relativity
CA-AstroPh	18771	198050	17903	196972	22.0	.669	14	4.77	Collaboration network, astrophysics
Gnutella09	8114	26013	8104	26008	6.42	.0137	10	5.22	Peer-to-peer filesharing network
EmailEnron	36692	183831	33696	180811	10.7	.708	13	4.72	E-mail network of Enron
Oregon1	11174	23409	11174	23409	4.19	.453	10	4.28	AS peering information
FB Graphs									
LehighFB	5075	198347	5073	198346	78.2	.270	6	3.19	Facebook friend network from Lehigh
VanderbiltFB	8096	427832	8063	427829	106	.255	7	3.18	Facebook friend network from Vanderbilt
StanfordFB	11621	568330	11586	568309	98.1	.252	9	3.35	Facebook friend network from Stanford
Miscellaneous Graphs									
PowerGrid	4941	6594	4941	6594	2.67	.107	46	24.2	Western US power grid
Polblogs	1224	16715	1222	16714	27.4	.360	8	3.43	Political blogs network
PlanarGrid	2500	4900	2500	4900	3.92	0.00	98	73.0	50-by-50 planar grid
RandGrid(3)	2500	3808	114	205	3.60	.510	34	21.4	Random planar graph, average degree 3
RandGrid(7)	2500	8679	2480	8656	6.98	.596	68	55.7	Random planar graph, average degree 7

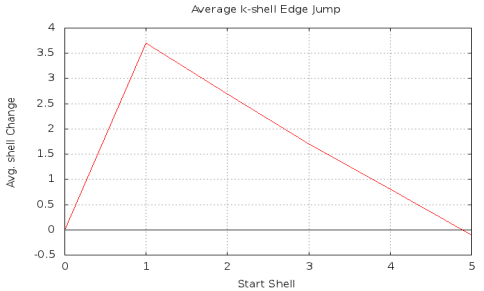
Some initial k-core results

Adcock, Sullivan, and Mahoney (2012)

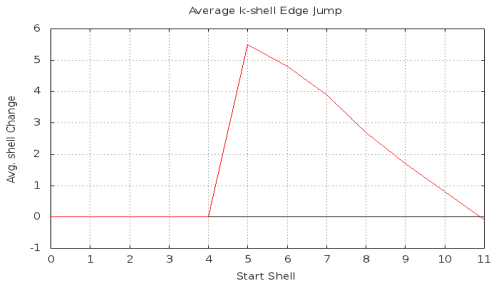
The effect of extreme sparsity in (real and model) random/noisy graphs: Erdos-Renyi, G_{np}, n=2500:



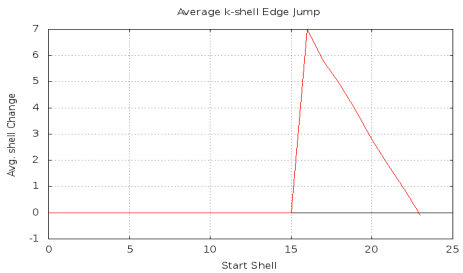
p=.0008, giant component



p=.0016, giant component



p=.0032

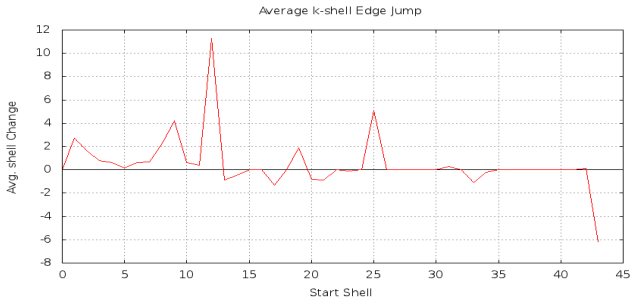


p=.0064

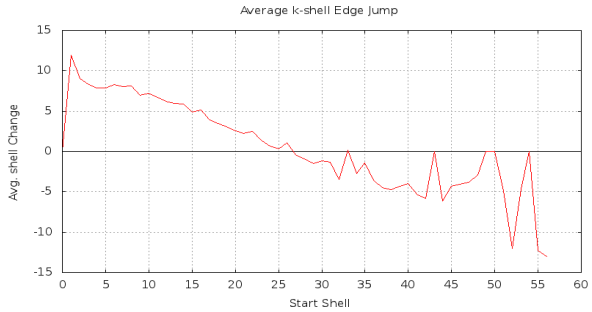
Some initial k-core results

Adcock, Sullivan, and Mahoney (2012)

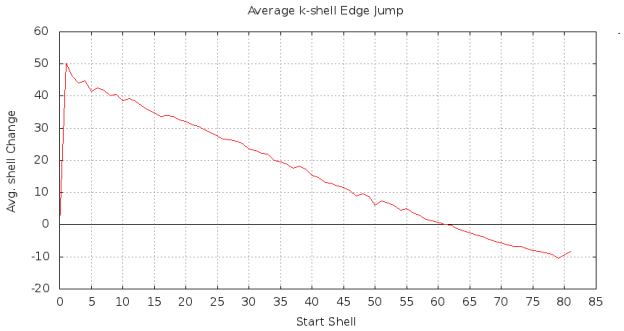
The effect of extreme sparsity in (real and model) random/noisy graphs:
Collaboration and social graphs:



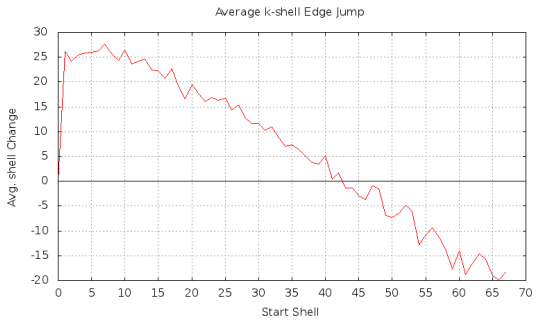
CA-GrQc, (~4K/26K nodes/edges)



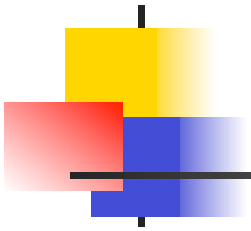
CA-AstroPhysics, (~18K/394K nodes/edges)



FB-Texas84, (~36K/3M nodes/edges)



soc-Epinions1, (~47K/730K nodes/ edges)



Some (*oversimplified*) summary stats…

Adcock, Sullivan, and Mahoney (2012)

ca-AstroPhysics:

- ~0.6% of nodes (113 nodes) in two deepest cores (k = 55,56)
- ~1.8% of edges (~7,000 edges) leaving the deepest core (k = 56)
- ~1.8% of edges (~7000 edges) leaving next core (k = 55)
- Max average k-shell change is +12 (out of k = 56 max shell)

Suggests collaborators tend to collaborate with people of similar coreness/peripheryness

"*Typical*" for collaboration graphs (and other NCP core-periphery graphs)

Texas84:

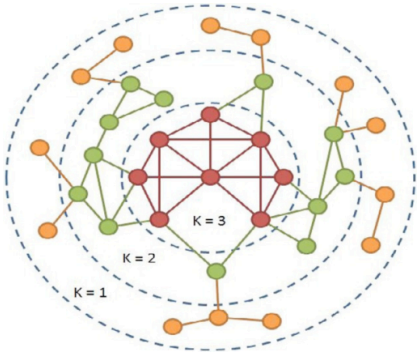
- ~8% of nodes (≥2400 nodes) in two deepest cores (k = 80,81)
- ~7% of edges (≥220K edges) leaving the deepest core (k = 81)
- ~17% of edges (≥510K edges) leaving the next core (k = 80)
- Max average k-shell change is +50 (out of k = 80 max shell)

Suggests that the "periphery" nodes are more tightly connected to "core-like" nodes

"*Typical*" for more social graphs (and Facebook in particular)

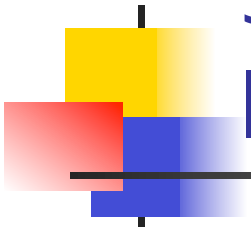
More k-core statistics

Adcock, Sullivan, and Mahoney (2014)



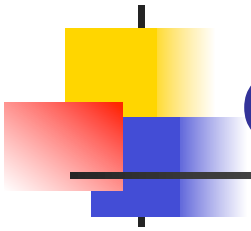
Network	n_c	$\bar{d}$	k_{min}	k_{max}	P_{kmin}	P_{kmax}
ER Networks						
ER(1.6)	3210	2.16	1	2	58.7	41.3
ER(1.8)	3617	2.28	1	2	49.6	50.4
ER(2)	4001	2.39	1	2	41.5	58.5
ER(4)	4879	4.05	1	3	8.28	67.6
ER(8)	4998	8.04	1	5	.300	88.3
ER(16)	5000	16.1	4	11	0.04	88.3
ER(32)	5000	32.1	7	23	0.02	93.8
PL Networks						
PL(2.25)	3393	3.32	1	5	45.7	0.825
PL(2.50)	4895	2.78	1	4	52.2	0.776
PL(2.75)	4650	2.43	1	2	59.0	41.0
PL(3.00)	4071	2.24	1	2	65.9	34.1
Planar Networks						
PlanarGrid	2500	3.92	2	2	100	100
PowerGrid	4941	2.67	1	5	32.1	0.243
Real Networks						
as20000102	6474	3.88	1	12	37.9	0.324
oregon1	11174	4.19	1	17	35.3	0.269
LehighFB	5073	78.2	1	62	1.42	15.4
VanderbiltFB	8063	106.1	1	86	1.98	23.3
StanfordFB	11586	98.1	1	91	4.36	20.1
CA-GrQc	4158	6.46	1	43	17.9	1.06
CA-AstroPh	17903	22.0	1	56	5.55	0.318
EmailEnron	33696	10.7	1	43	28.4	0.816
EmailEuAll	224832	3.02	1	37	83.9	0.130
Polblogs	1222	27.4	1	36	11.3	4.50
StanfordWeb	255265	15.2	1	71	5.98	0.152

Table 1: k-core network statistics: P_{kmin} and P_{kmax} are percentage of nodes in the k_{min} and k_{max} shells, respectively.



Summary of results for notions of tree-like-ness & core-periphery

- δ hyperbolicity and tree decompositions in general capture very different manners in which graphs can be tree-like
- (Good since NCP results of LLDM08 show realistic social graphs are very different in terms of both metric and cut structure)
- For realistic social graphs, they often capture very similar core-periphery structures, also captured with k-core decompositions
- Different tree decomposition heuristics behave in characteristic ways on toy graphs
- "Extremely sparse" versus "very sparse" leads to different k-core (and many other) properties
- k-core properties of real graphs differ a lot, e.g., number of deep cores and number of nodes in deep cores



Outline

Background & failure of simple statistics

- Intermediate-scale structure

Notions of tree-like-ness & general thoughts on core-periphery

- Metric-based versus cut-based versus k-core-based notions of tree-like-ness

δ-hyperbolicity in practice and in theory

- Many real graphs have size-resolved hyperbolic properties
- Too much randomness in models for theory to be nontrivial

Tree decompositions

- Many practical heuristics developed in scientific computing
- Tool for identifying small-scale and large-scale structure in real graphs

Connections with prior work

- On core-periphery & size-resolved isoperimetric/clustering/community structure

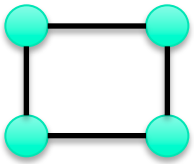
δ_d: scaled and size resolved δ

$$\delta_d = \max_{\phi(x,y,u,v)=d} \frac{\delta(x,y,u,v)}{d}$$

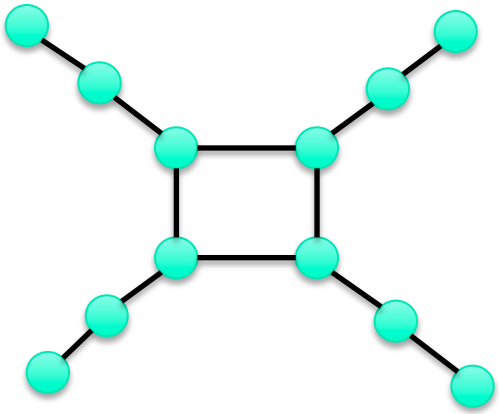
Let $\varphi(x,y,u,v)$ be the maximum pairwise distance between the four points.

Then, δ_d gives the hyperbolicity of structures of size d as a function of d.

Example:



δ = 1.0



δ = 1.0

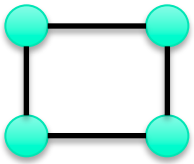
δ_d: scaled and size resolved δ

$$\delta_d = \max_{\phi(x,y,u,v)=d} \frac{\delta(x,y,u,v)}{d}$$

Let $\varphi(x,y,u,v)$ be the maximum pairwise distance between the four points.

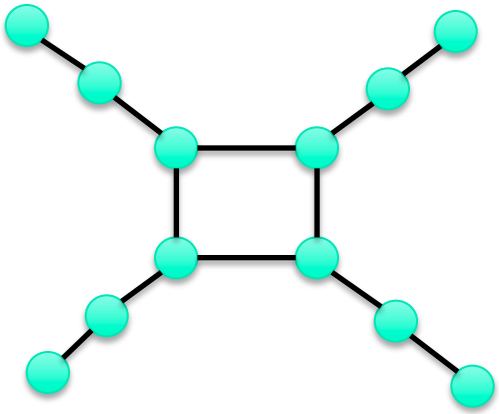
Then, δ_d gives the hyperbolicity of structures of size d as a function of d.

Example:



δ = 1.0

δ_2 = 0.5

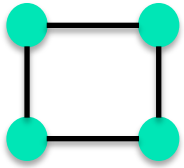


δ = 1.0

δ_2 = 0.5

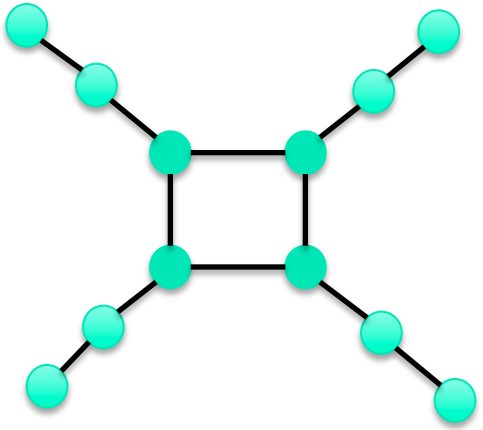
δ_6 = 0.16

Example plot



$\delta = 1.0$

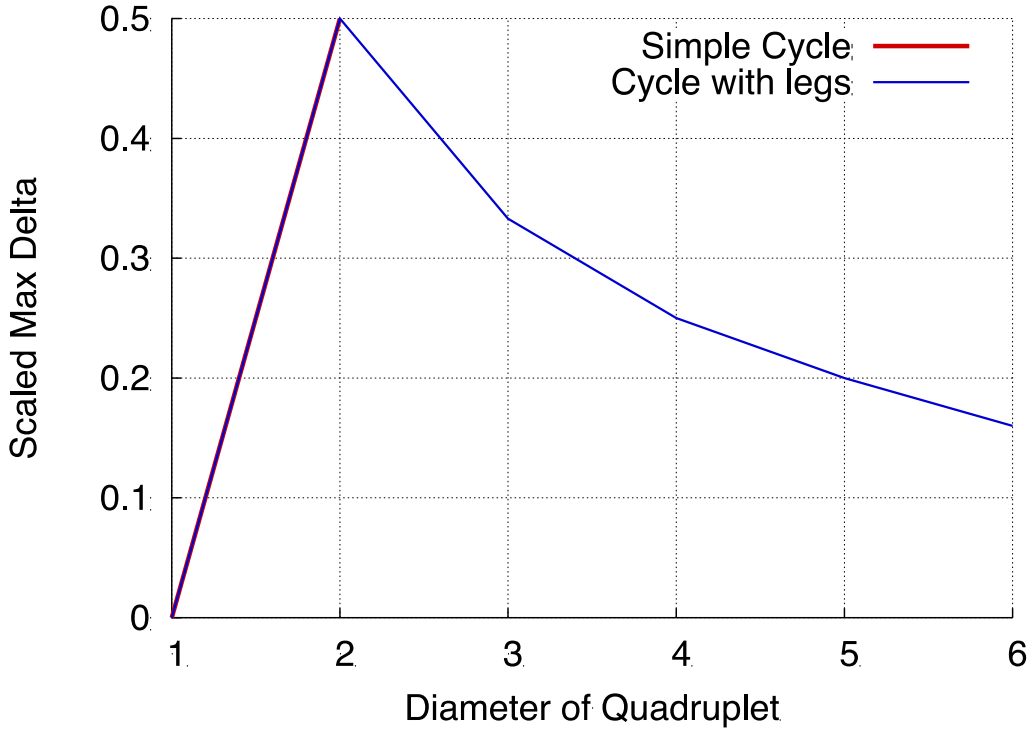
$\delta_2 = 0.5$



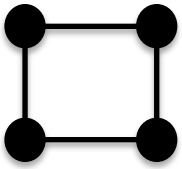
$\delta = 1.0$

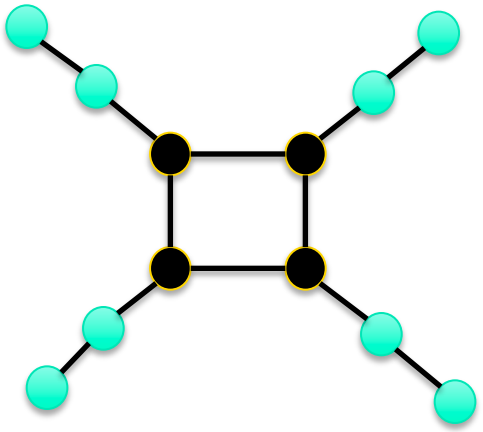
$\delta_2 = 0.5$

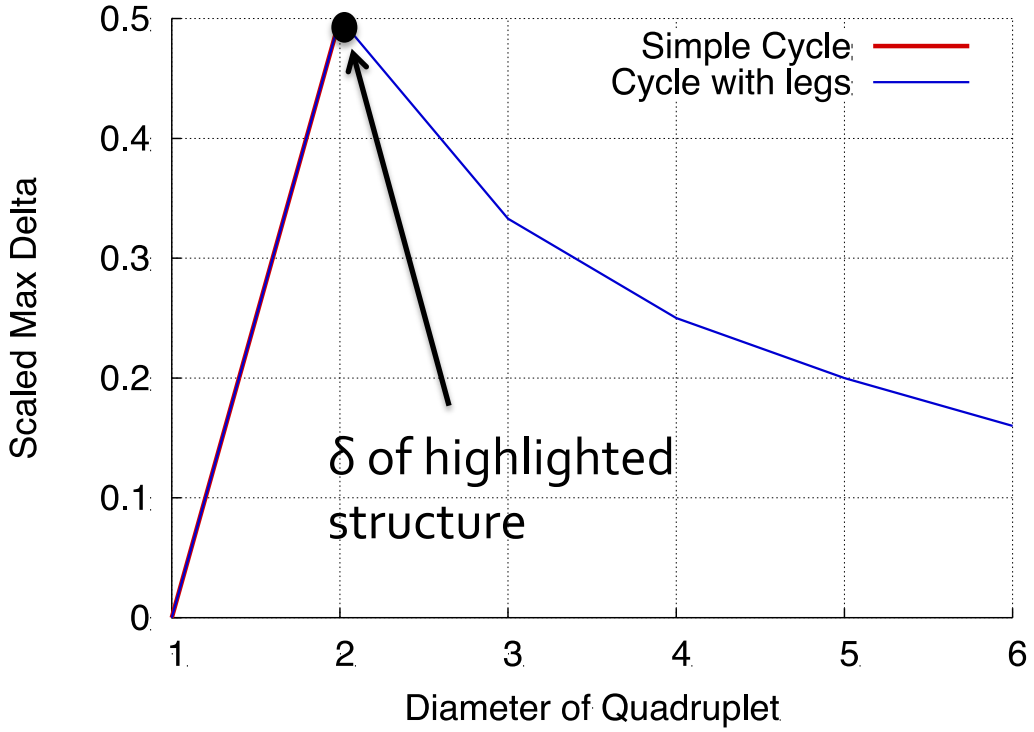
$\delta_6 = 0.16$



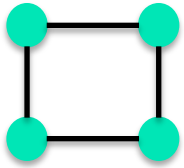
Example plot

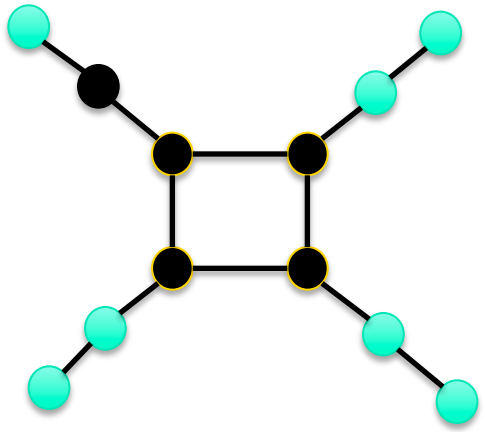


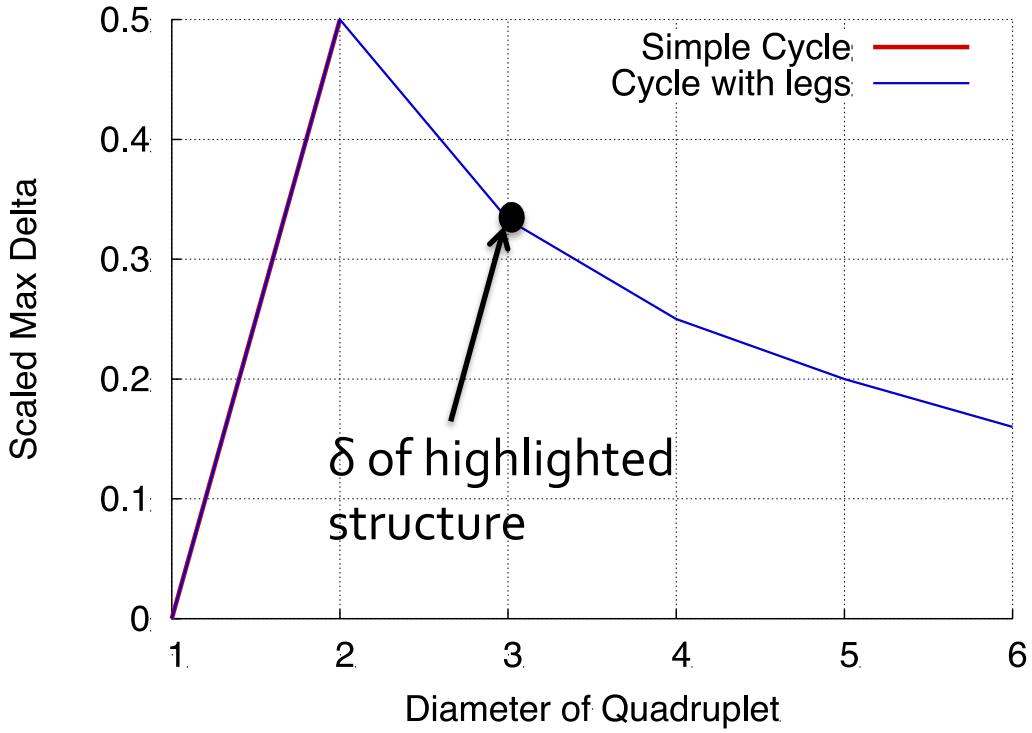




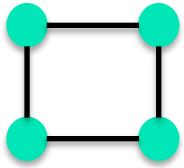
Example plot

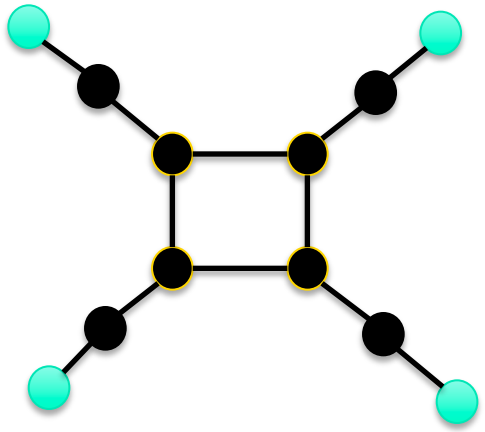


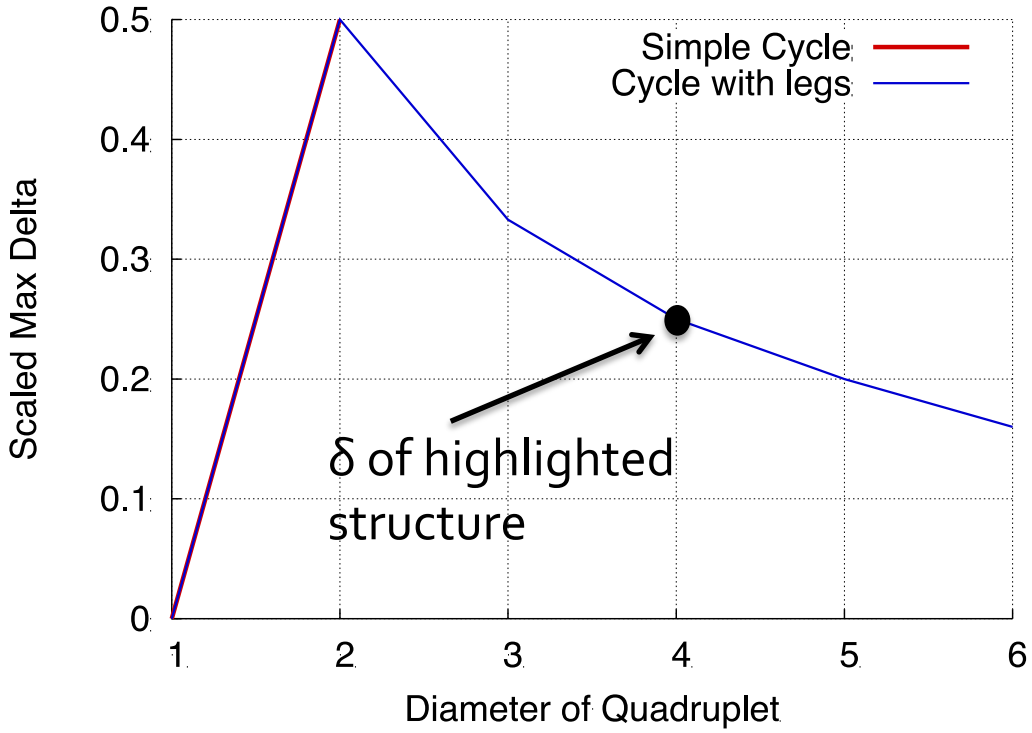




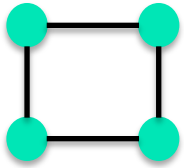
Example plot

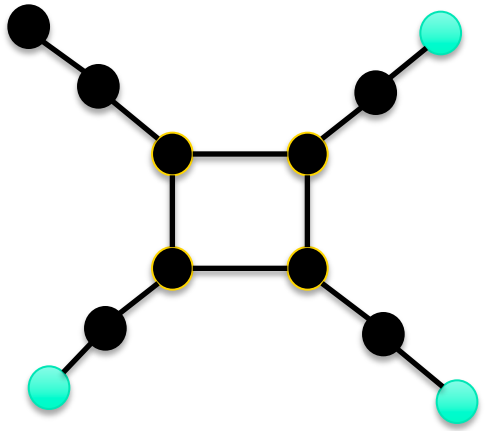


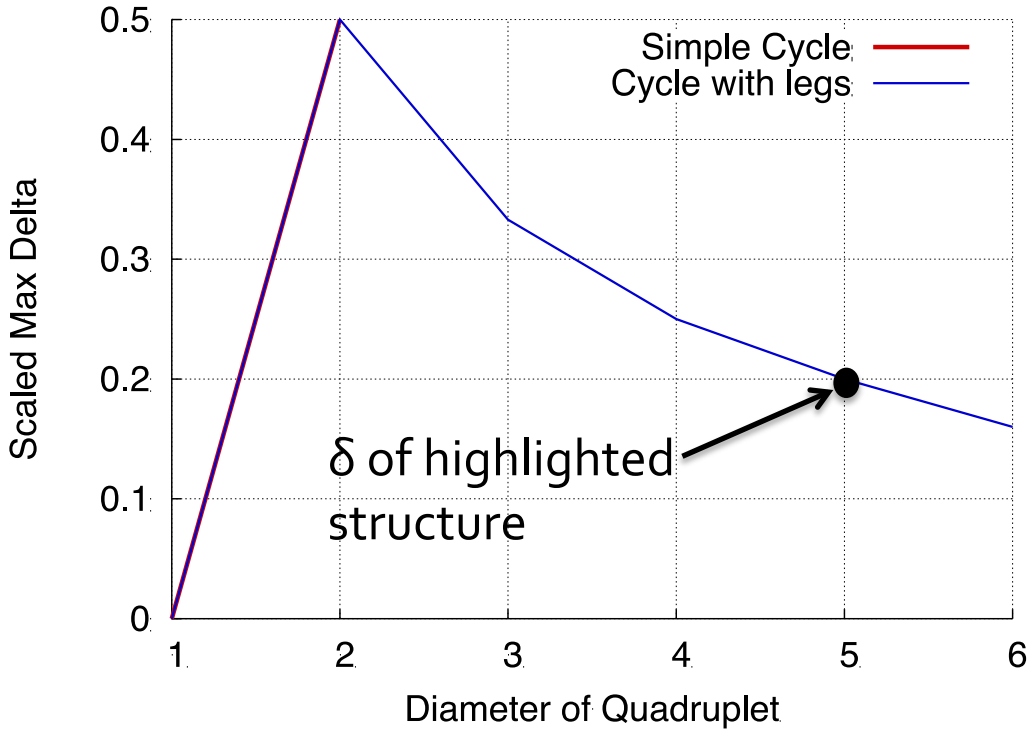




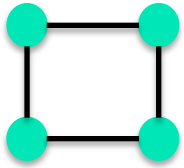
Example plot

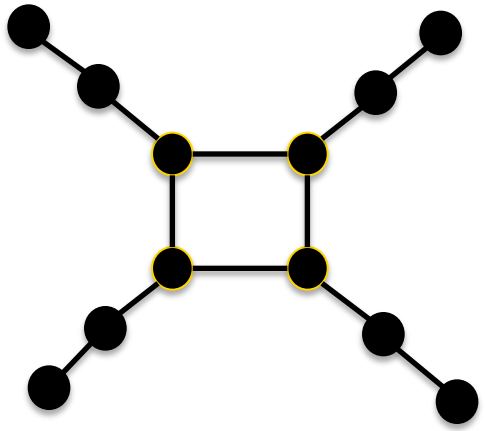


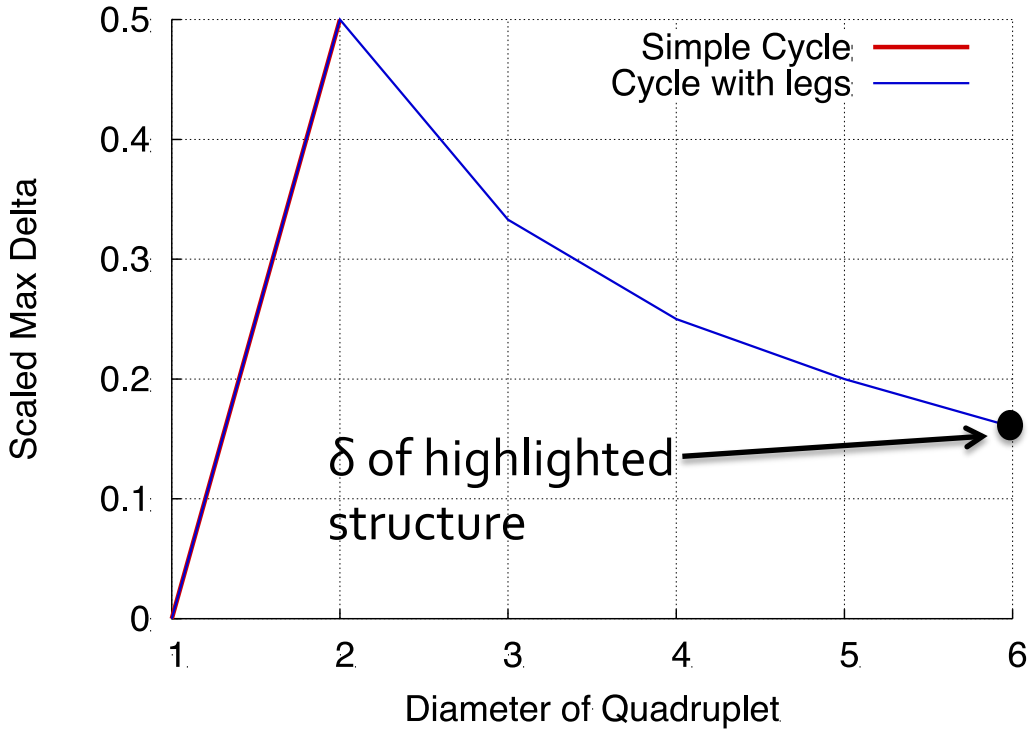




Example plot

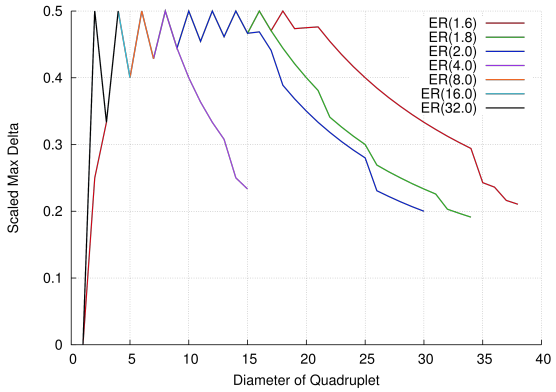




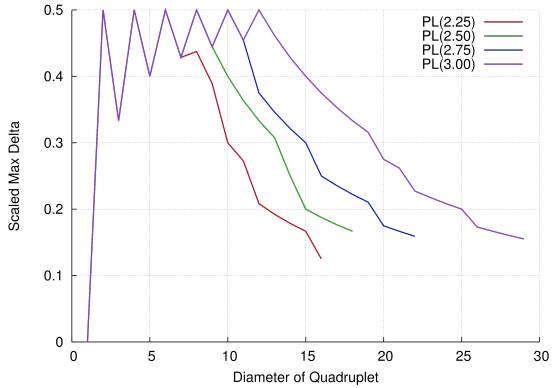


Max δ_D versus diameter of quadruplet

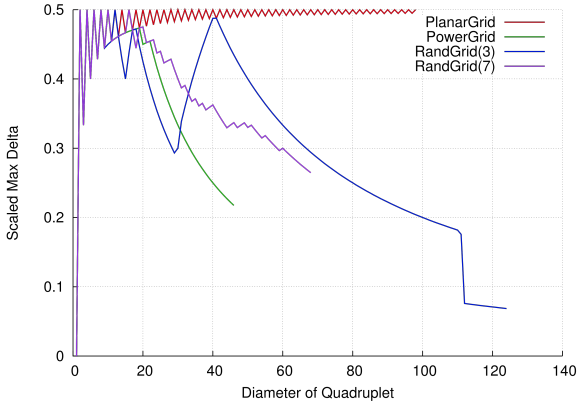
Adcock, Sullivan, and Mahoney (2014)



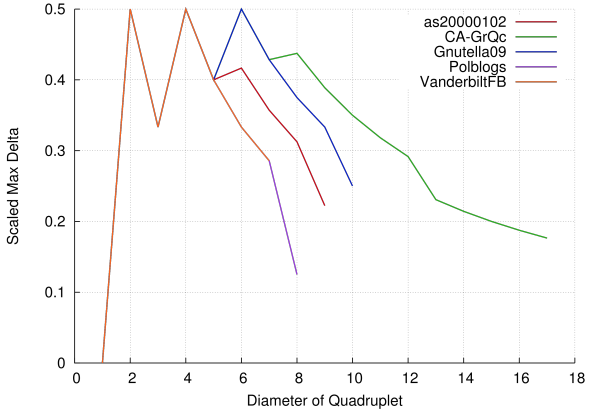
ER graphs



PL graphs



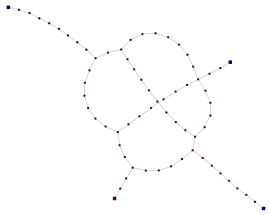
"Planar" graphs



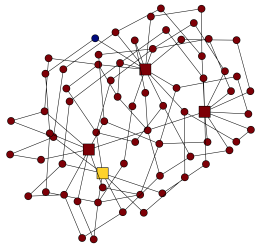
Real social graphs

What's driving hyperbolicity? Selected quadruplets, with all geodesics shown

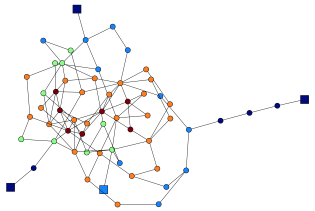
Adcock, Sullivan, and Mahoney (2014)



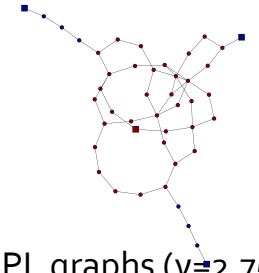
ER graph (p=1.6)



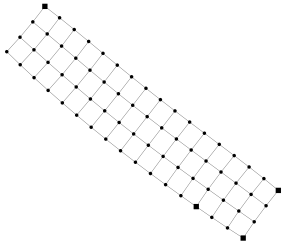
ER graphs (p=32)



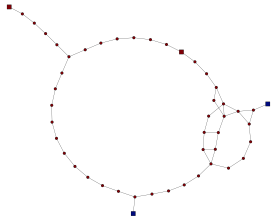
PL graphs (γ=2.25)



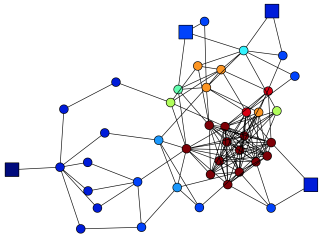
PL graphs (γ=2.75)



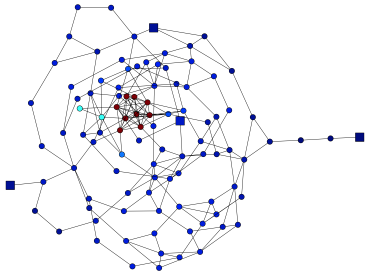
PlanarGrid



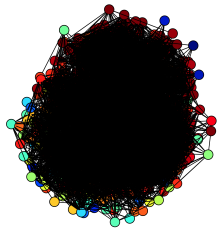
US PowerGrid



as20000102

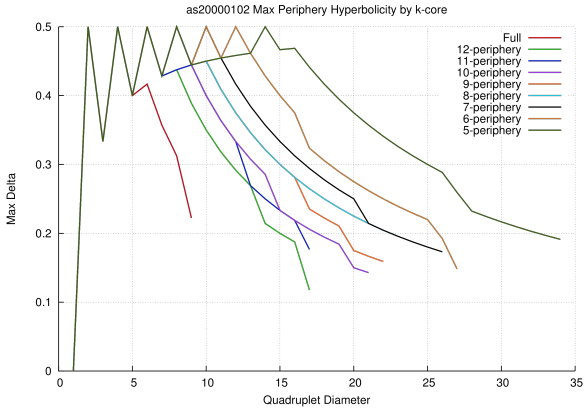


CA-GrQc

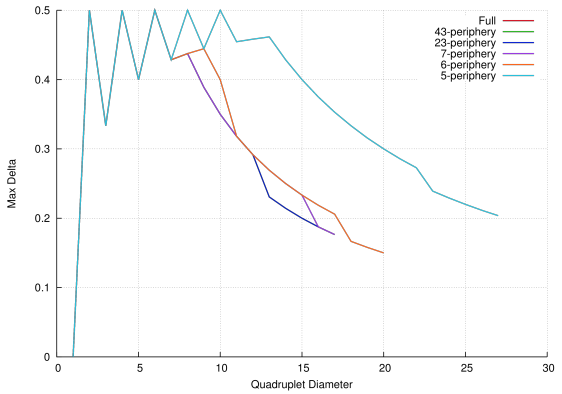


VanderbiltFB

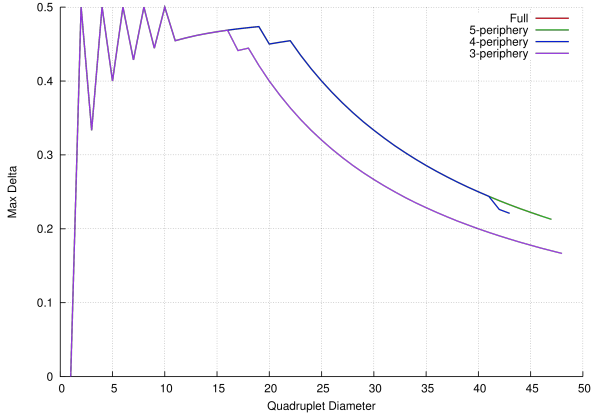
Maximum hyperbolicity of network periphery



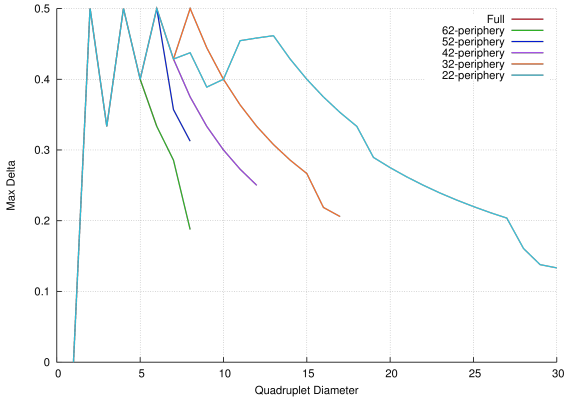
as20000102



CA-GrQc



PowerGrid



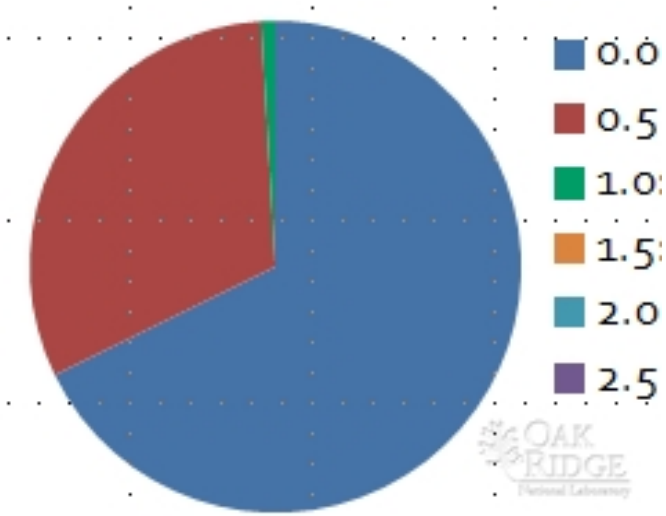
LehighFB

Computing δ: Sampling versus Brute Force

Adcock, Sullivan, Hernandez, and Mahoney (2013)

- Computing δ takes $O(n^4)$ time; prior works sampled to estimate δ.
- E.g., sample about **.0002 percent** quadruplets; although biased towards pairs at larger distances, this could still *easily miss the maximum δ*.
- It is *likely ok for computing average δs* (but it's not clear if that is useful).
- Example below is SNAP graph as20000101 (about 1600 nodes); max delta is achieved on **2×10^{-11} percent** of quadruplets.
- Took $>10^5$ CPU hours on Oak Ridge's NICS Nautilus supercomputer (1016 cores, 4TB shared memory) stress testing of OpenMP workshare and tasking models to parallelize computations at scale

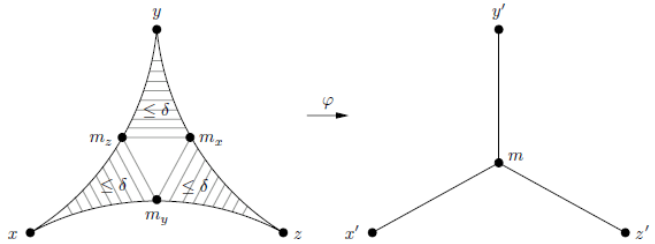
δ	*Fraction of quadruplets:*	*# of quadruplets*
0.0:	0.677473774788751	4577453756970
0.5:	0.313235924997126	2116425779202
1.0:	0.009262044976055	62580404070
1.5:	0.000028008357243	189242691
2.0:	0.000000246259522	1663890
2.5:	**0.000000000022835**	**154**
Total	0.999999999401533	6756650846976



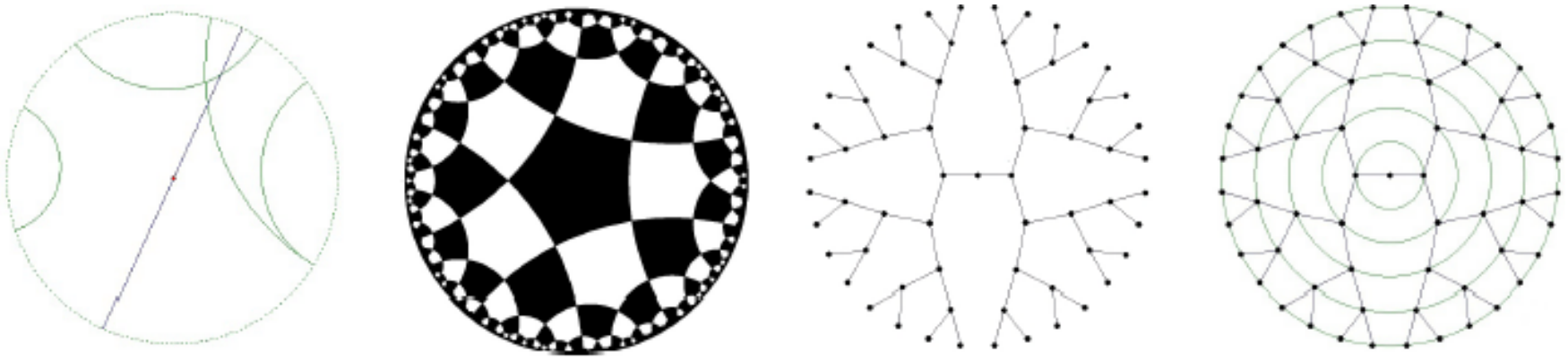
Another view on δ-hyperbolicity

Chen, Feng, Hu, and Mahoney (2012)

- Viewing graphs as geodesic metric spaces (replace edges with length 1 segments intersecting only at endpoints) provides another way to think of δ-hyperbolicity.
- For a geodesic triangle, there is a unique isometry to a "tripod" so that except for the leaves , each point on the tripod has two pre-images on the triangle.
- A triangle is **δ-thin** if the pre-images of every tripod point have distance at most δ.



- Classify graphs based on their δ-hyperbolicity relative to the graph diameter (similar to Jonckheere et al). Say a class is:
  - **constantly** hyperbolic if their δ's are constant, regardless of the size or diameter;
  - **logarithmically** hyperbolic if their δ's are O(log diameter);
  - **not** hyperbolic if their δ's are at the same order as the graph diameters.



Chen, Feng, Hu, and Mahoney (2012)

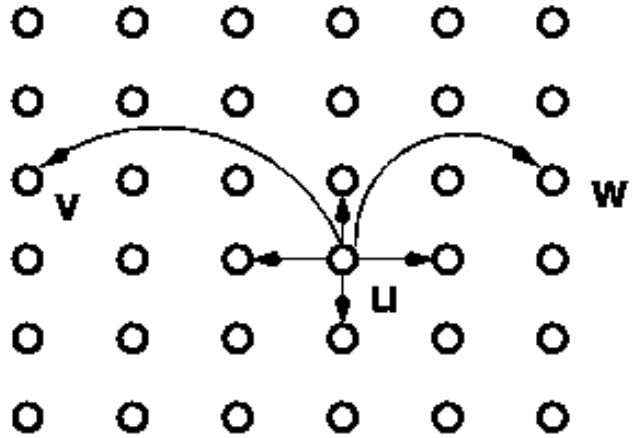
- Definition: A **ringed tree** is a binary tree with extra edges added that connect all vertices at a given tree level into a ring.

- Theorem [CFHM12]: This ringed tree model is quasi-isometric to the Poincare disk, and thus has constant hyperbolicity.
- Theorem [CFHM12]: If long-range edges between the leaves of a ringed tree are added according to a probability function that decreases:
  - exponentially fast with the ring distance, then we get logarithmically hyperbolic random graphs
  - as a power-law with the ring distance, then we get non-hyperbolic random graphs
- Theorem [CFHM12]: If one replaces the ringed tree with a pure binary tree, none of the resulting graphs are hyperbolic.

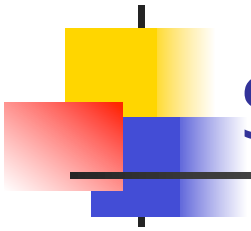
Example: Small world graphs

Chen, Feng, Hu, and Mahoney (2012)

- **Kleinberg's small-world random graphs:** start with a d-dimensional grid; add long-range edges between vertices u, v with probability proportional to $1/d_B(u,v)^p$, where p is a parameter of the model.
- Theorem [Kle00]: Efficient decentralized navigation is not attainable unless $p = d$.

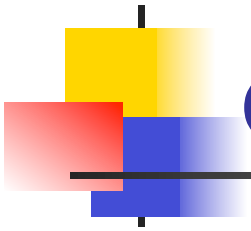


- Theorem [CFHM12]: Even at the "sweetspot" of $p = d$, with high probability, the small-world graphs are not logarithmically hyperbolic.
- Theorem [CFHM12]: When $p < d$, the small-world graphs are not hyperbolic; while when $p > 3$ and $d = 1$, the hyperbolic delta is polynomial in the size of graph and thus is also not logarithmically hyperbolic.
- **The point:** Long-range edges that enable efficient navigation do not significantly improve the hyperbolicity of the graphs (relative to their diameter).



Summary of results for δ hyperbolicity

- δ is expensive, coarse, and brittle
- Scaled and size-resolved δ reveals small-scale cyclic structures and large-scale hyperbolic properties in many real graphs
- Constant-degree expanders and well-formed meshes are *not* hyperbolic, relative to their diameter
- Even a very little structural heterogeneity enhances hyperbolicity
- Extremely sparse random graphs have very different hyperbolic properties than only very sparse random graphs
- There is "too much" randomness in existing network generative models to capture δ hyperbolicity properties



Outline

Background & failure of simple statistics

- Intermediate-scale structure

Notions of tree-like-ness & general thoughts on core-periphery

- Metric-based versus cut-based versus k-core-based notions of tree-like-ness

δ-hyperbolicity in practice and in theory

- Many real graphs have size-resolved hyperbolic properties
- Too much randomness in models for theory to be nontrivial

Tree decompositions

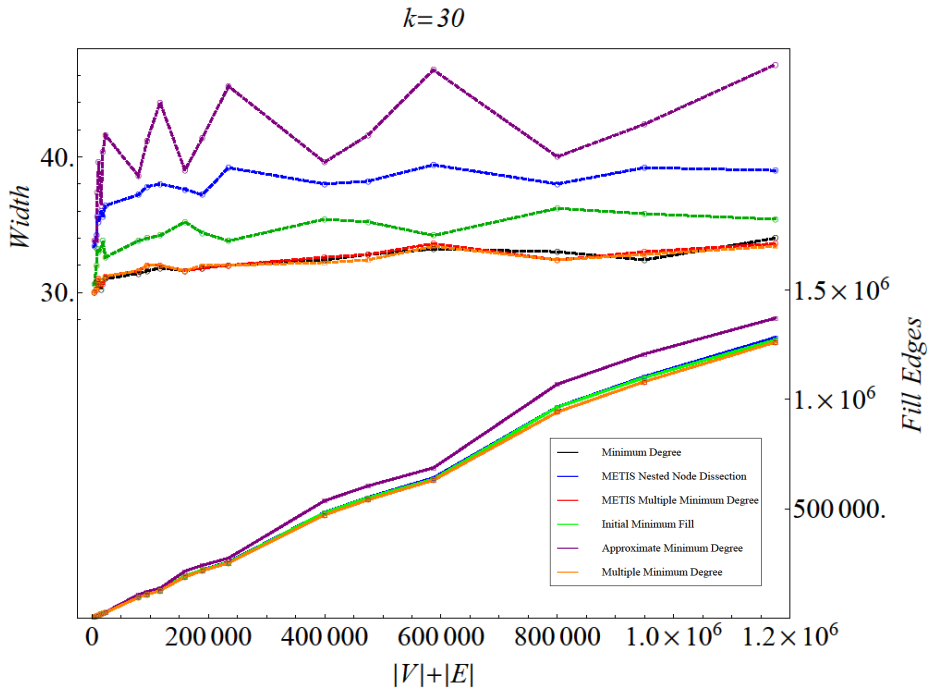
- Many practical heuristics developed in scientific computing
- Tool for identifying small-scale and large-scale structure in real graphs

Connections with prior work

- On core-periphery & size-resolved isoperimetric/clustering/community structure

Prior uses of tree decompositions

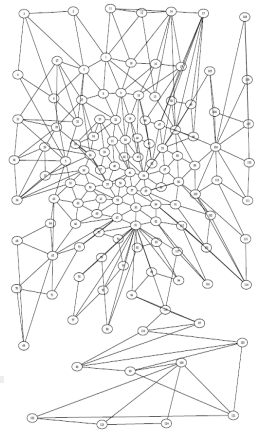
- In numerical linear algebra: one often wants to permute the rows of a matrix before computing a factorization so that the resulting factors are as sparse as possible, so minimize the number of "fill edges" added.



Comparison of width and fill from 6 heuristics on graphs known to have tw <= 30

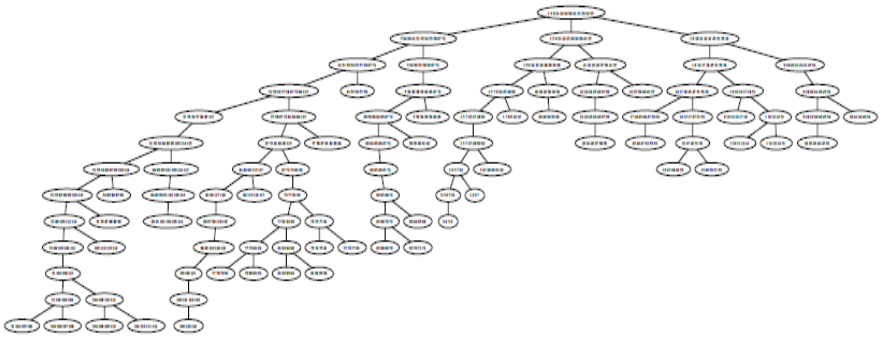
- For tree decompositions, we instead need to minimize the maximum clique size in the resulting chordal graph.
- Numerous implementations of common heuristics are available, and we tested several on a large set of random graphs with a fixed maximum width and varying sizes.
- Min-degree-based heuristics are orders of magnitude faster than min-fill, etc.

Problems with tree decompositions



Adcock, Sullivan, and Mahoney (2012)

- Every bag in a tree decomposition is a vertex separator, so a low-width decomposition means many small separators.
- Bit discrepancies between upper bounds given by min degree/min fill heuristics and MMD lower bounds
- Treewidth is O(n) w/ high probability for many random graphs (Gao 2009):
  - Erdos-Renyi graphs G(n,m), *when m/n > 1.073.*
  - Random intersection graphs G(n,m,p) on {1,…m}, *with $m=n^a$, p at least 2/m and a > 0.*
  - Preferential attachment (BA) graphs, *with at least 12 new edges for each additional vertex.*
- Current heuristics optimize treewidth/treelength – but maybe 1 bad bag (e.g., consisting of high degree nodes) is OK.
- Current heuristics get lost in "local noise"

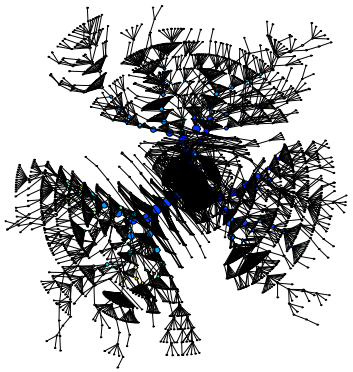


Examples of (AMD) tree decompositions

Adcock, Sullivan, and Mahoney (2014)

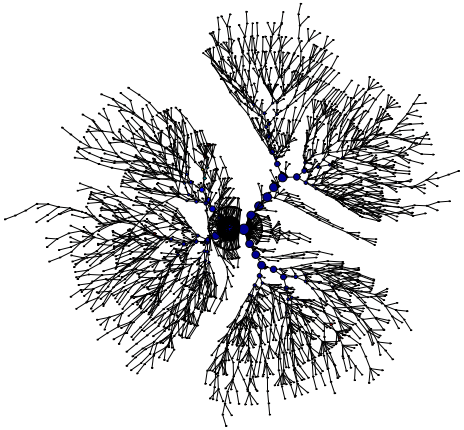
as2000102

Nmbr of nodes: 6474
Treewidth: 88



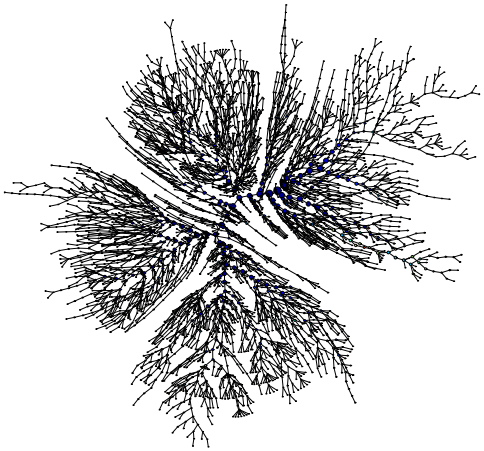
CA-GrQc

Nmbr of nodes: 4158
Treewidth: 222



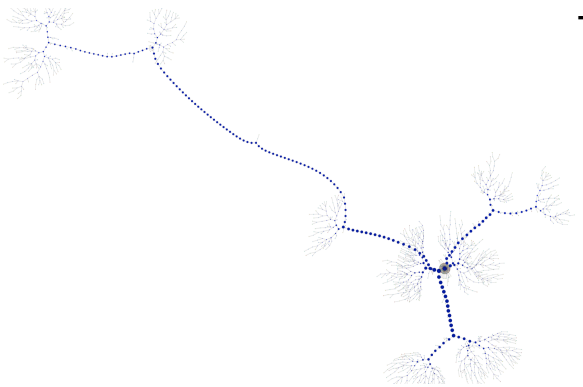
PowerGrid

Nmbr of nodes: 4941
Treewidth: 21



FB-Lehigh

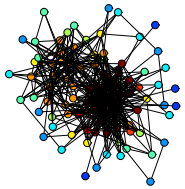
Nmbr of nodes: 5073
Treewidth: 2983



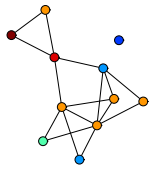
What do bags look like in real graphs?

Adcock, Sullivan, and Mahoney (2014)

as2000102



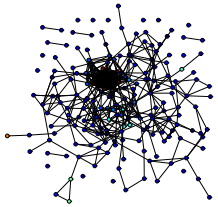
Core



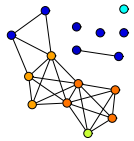
Interm.

Periphery

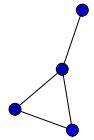
CA-GrQc



Core

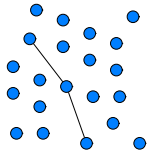


Interm.



Periphery

PowerGrid

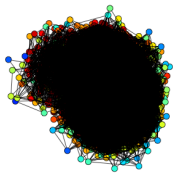


Core

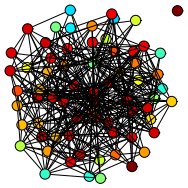
Interm.

Periphery

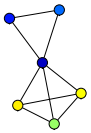
FB-Lehigh



Core



Interm.

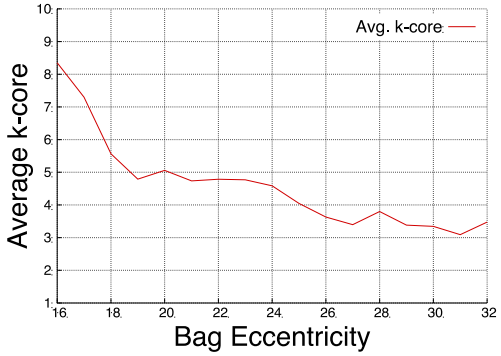


Periphery

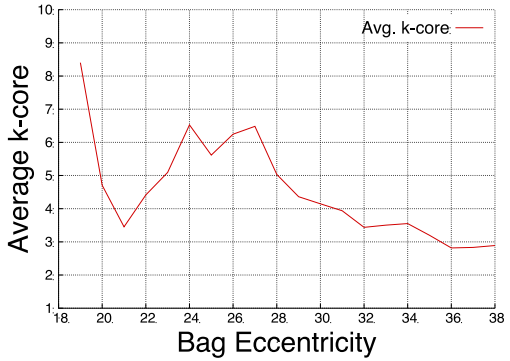
K-core versus eccentricity

Adcock, Sullivan, and Mahoney (2014)

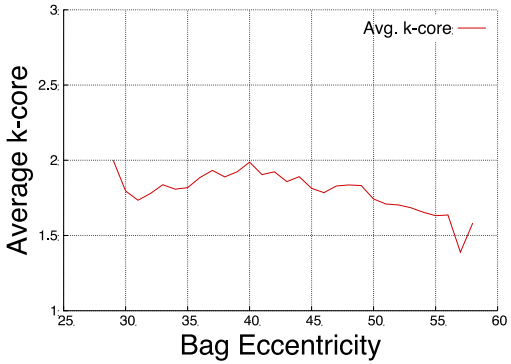
as2000102



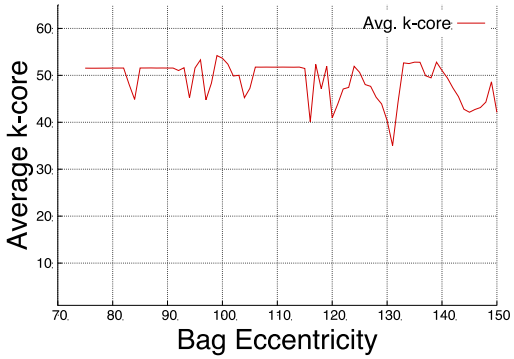
CA-GrQc



PowerGrid



FB-Lehigh

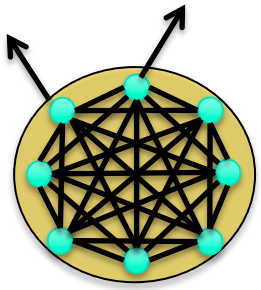


Tree decompositions and good-conductance communities

"It is a matter of common experience that communities exist in networks … Although not precisely defined, communities are usually thought of as sets of nodes with better connections amongst its members than with the rest of the world."

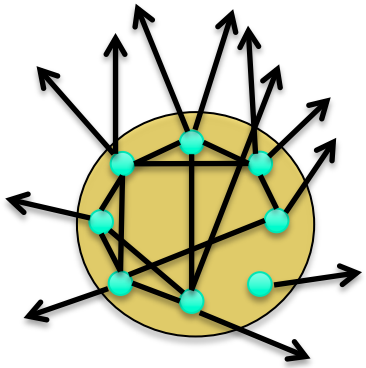
Examples of conductance (defn)

Edges to rest of network



Low conductance (good) community

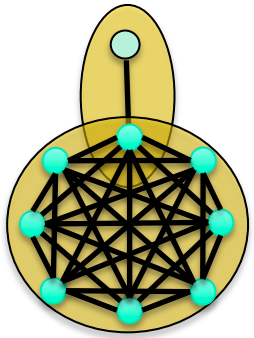
Edges to rest of network



High conductance (bad) community

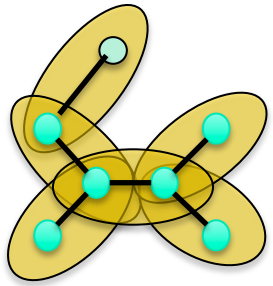
Examples of conductance (with TDs)

Rest of network



Number of bags to cover: 2

Rest of network

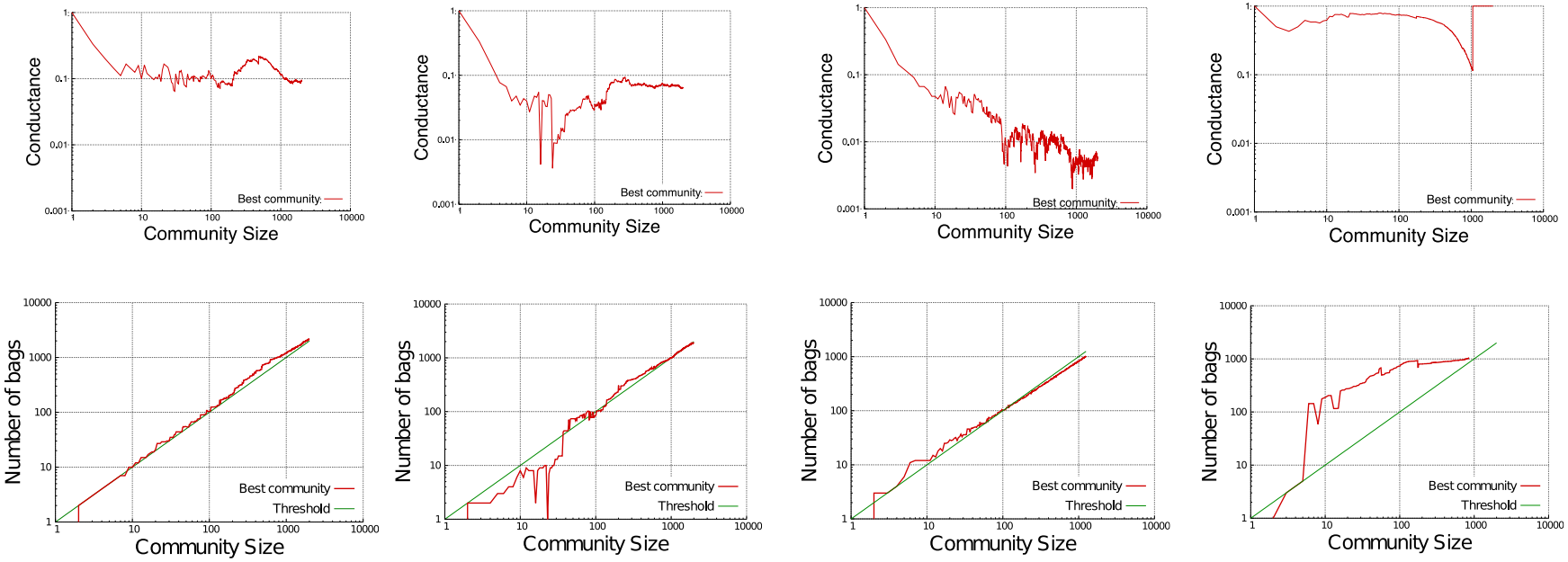


Number of bags to cover: n

Tree decompositions and good-conductance communities

Adcock, Sullivan, and Mahoney (2014)

as2000102 | CA-GrQc | PowerGrid | FB-Lehigh



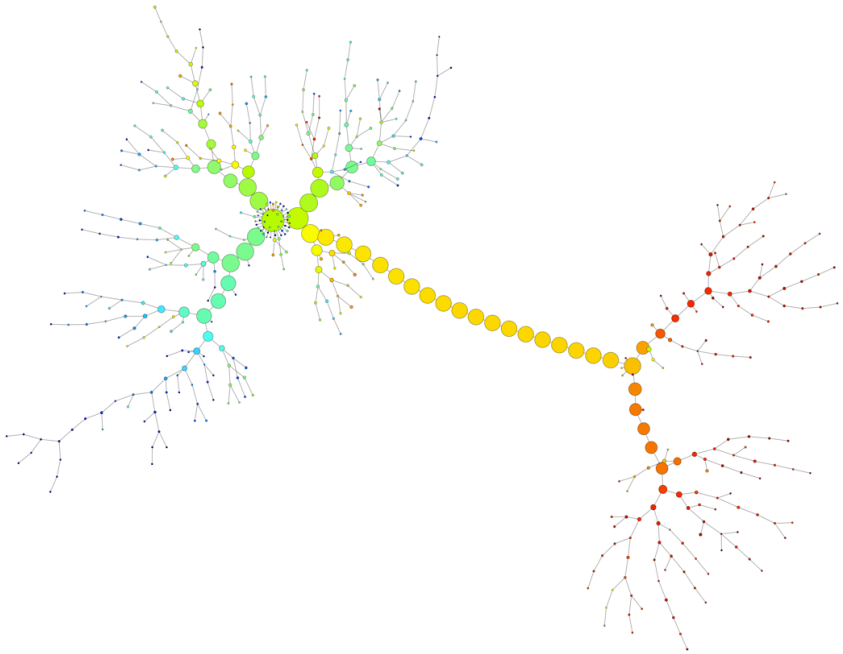
- *When small-cardinality good-conductance clusters exist, they are well-localized in the TD (CA-GrQc)*
- *When small-cardinality bad-conductance clusters are present, they are poorly-localized in the TD (FB-Lehigh)*

Tree decompositions and bad-conductance communities

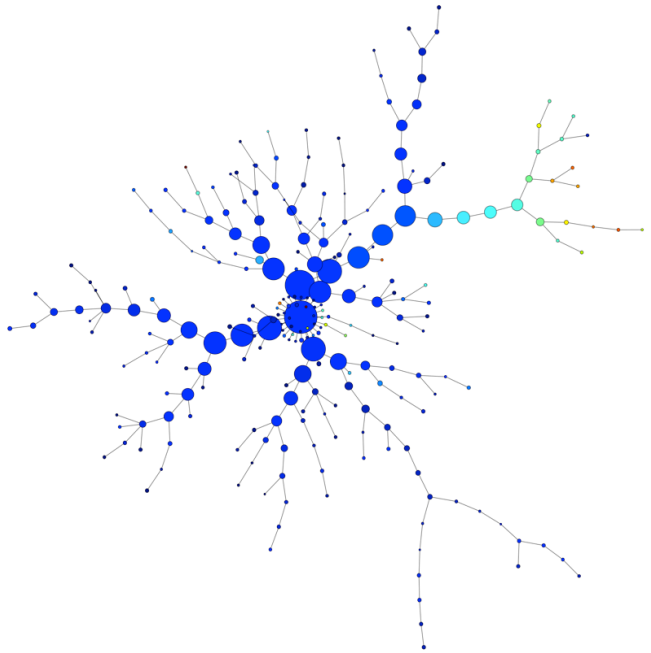
Adcock, Sullivan, and Mahoney (2014)

"Communities are just some set of nodes that are similar in a way that may or may not be at all related to the graph."

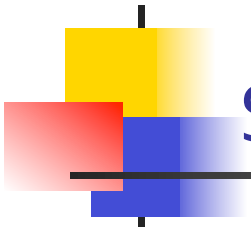
Consider communities defined by two types of meta-information (graduation year and residence):



FB-Haverford:
Bags colored by graduation year
(red = 2009, blue = pre-2004)



FB-Caltech:
Bags colored by residence (170)



Summary of results for tree decompositions

Previous work says real social graphs are *not* tree-like by TD measures

- basically since tree-width is large, but restricted internet router and AS networks

Existing TD heuristics are *not* well-suited to social graphs, but they can be used

- Open problem: develop TD heuristics more appropriate for social graphs

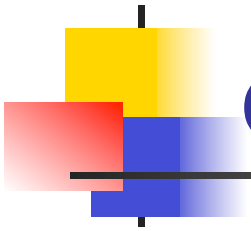
Tree-like core-periphery structure: shown by bag cardinality histograms, mean bag density versus bag cardinality, & mean k-core versus bag eccentricity plots

- that correlates with δ hyperbolicity and k-core structure
- captures larges-scale cycles and small-scale clusters

But there are some high-treewidth bags at deep cores

- tree-like-ness enhanced ad deep cores are removed
- for very social graphs, a huge fraction of nodes are in deep cores

Good-conductance and bad-conductance communities can be identified



Outline

Background & failure of simple statistics

- Intermediate-scale structure

Notions of tree-like-ness & general thoughts on core-periphery

- Metric-based versus cut-based versus k-core-based notions of tree-like-ness

δ-hyperbolicity in practice and in theory

- Many real graphs have size-resolved hyperbolic properties
- Too much randomness in models for theory to be nontrivial

Tree decompositions

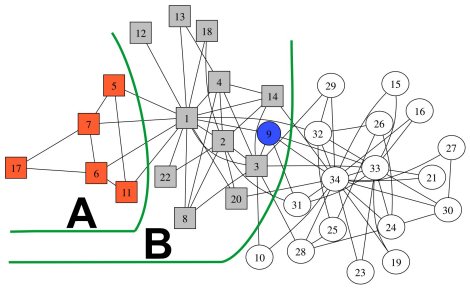
- Many practical heuristics developed in scientific computing
- Tool for identifying small-scale and large-scale structure in real graphs

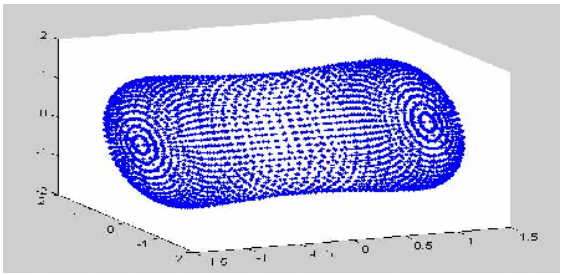
Connections with prior work

- On core-periphery & size-resolved isoperimetric/clustering/community structure

What do the data "look like" (if you *squint* at them)?

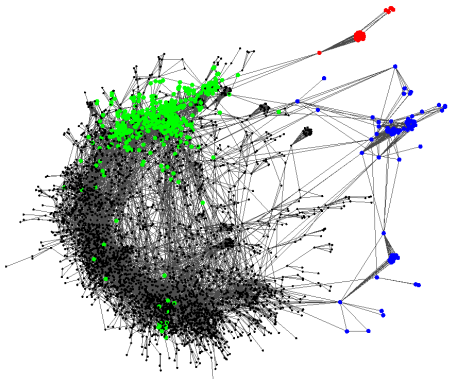
A "hot dog"?

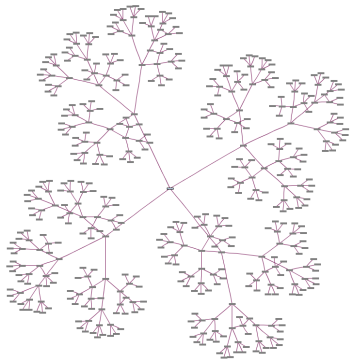




(or pancake that embeds well in low dimensions)

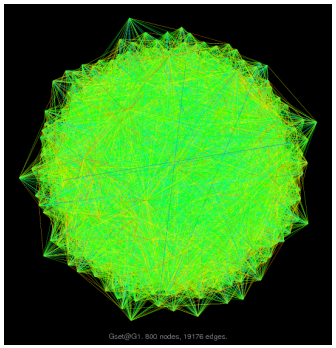
A "tree"?

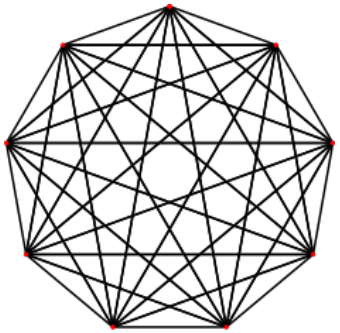




(or tree-like hyperbolic structure)

A "point"?





(or clique-like or expander-like structure)

Squint at the data graph …

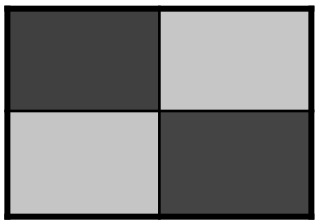
Say we want to find a "best fit" of the adjacency matrix to:

α	β
β	γ

What does the data "look like"? How big are α, β, γ?

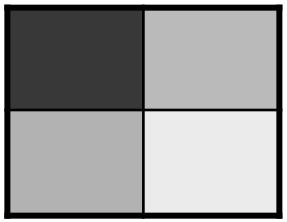
$\alpha \approx \gamma \gg \beta$

low-dimensional



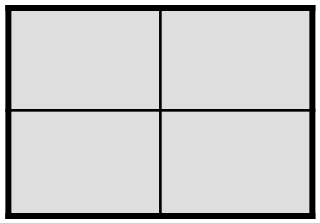
$\alpha \gg \beta \gg \gamma$

core-periphery



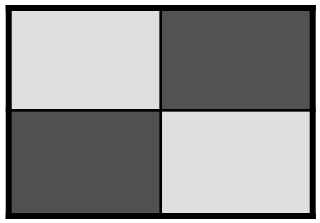
$\alpha \approx \beta \approx \gamma$

expander or K_n

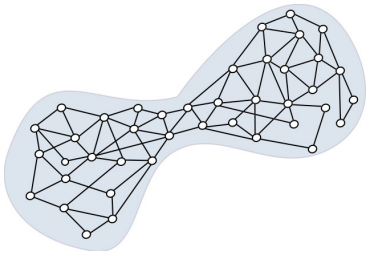


$\beta \gg \alpha \approx \gamma$

bipartite graph



Communities, Conductance, and NCPPs



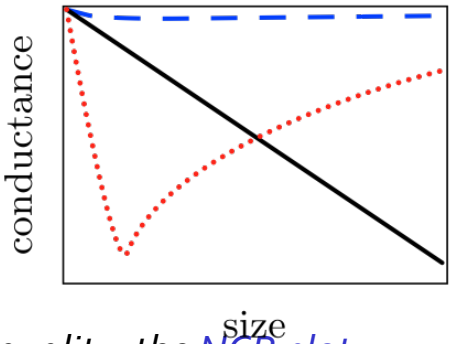
Let A be the adjacency matrix of G=(V,E).

The conductance ϕ of a set S of nodes is:

$$\phi(S) = \frac{\sum_{i \in S, j \notin S} A_{ij}}{\min\{A(S), A(\overline{S})\}} \qquad A(S) = \sum_{i \in S} \sum_{j \in V} A_{ij}$$

The Network Community Profile (NCP) Plot of the graph is:

$$\Phi(k) = \min_{S \subset V, |S| = k} \phi(S)$$



Just as conductance captures the "gestalt" notion of cluster/community quality, the NCP plot measures cluster/community quality ***as a function of size.***

NCP is intractable to compute --> use approximation algorithms!

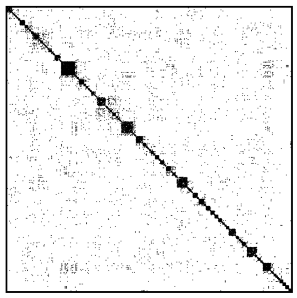
Since algorithms often have non-obvious size-dependent behavior.

NCPs and three types of graphs

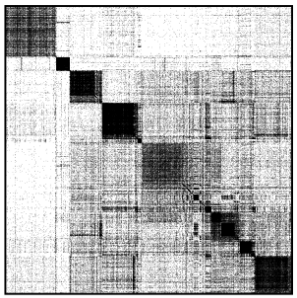
Jeub, Balachandran, Porter, Mucha, and Mahoney (2014)

	Nodes	Edges	$\langle k \rangle$	λ_2	$\langle C \rangle$	Description
CA-GrQc	4 158	13 422	6.5	0.0019	0.56	Coauthorship: `arXiv` general relativity
CA-AstroPh	17 903	196 972	22.0	0.0063	0.63	Coauthorship: `arXiv` astrophysics
FB-Johns55	5 157	186 572	72.4	0.1258	0.27	Johns Hopkins Facebook network
FB-Harvard1	15 086	824 595	109.3	0.0094	0.21	Harvard Facebook network
US-Senate	8 974	422 335	60.3	0.0013	0.50	Network of voting patterns in U.S. Senate
US-House	36 646	6 930 858	240.5	0.0002	0.58	Network of voting patterns in U.S. House

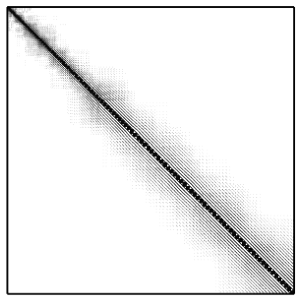
Table 1: Six medium-sized networks. For each network, we show the number of nodes and edges in the largest connected component (LCC), the mean degree/strength ($\langle k_i \rangle$), the second-smallest eigenvalue (λ_2) of the normalized Laplacian matrix, the mean clustering coefficient ($\langle C_i \rangle$), and a description.



CA-GrQc



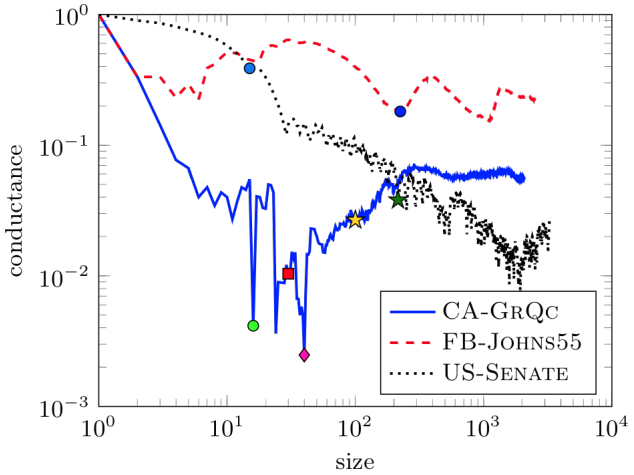
FB-Johns55

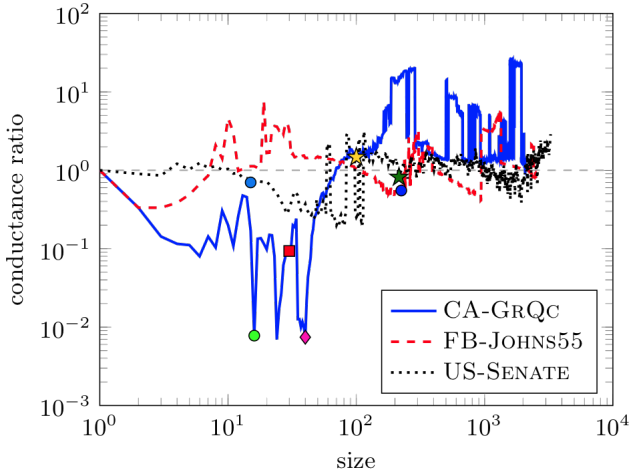


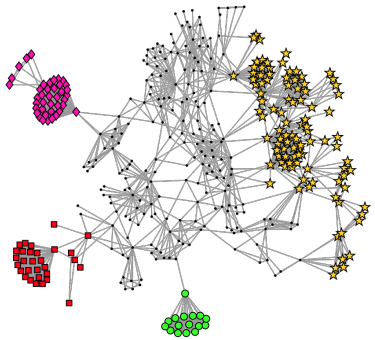
US-Senate

NCPs and core-periphery (or not)

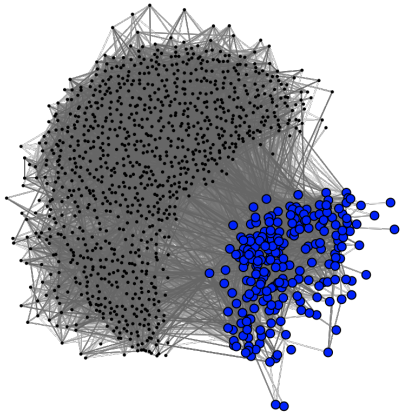
Jeub, Balachandran, Porter, Mucha, and Mahoney (2014)



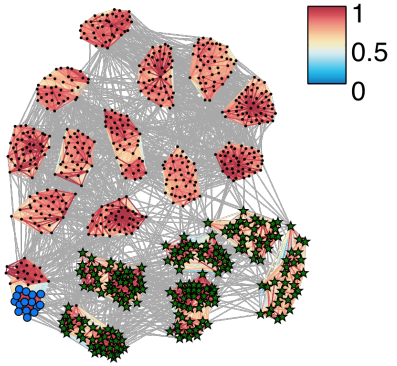




CA-GrQc



FB-Johns55

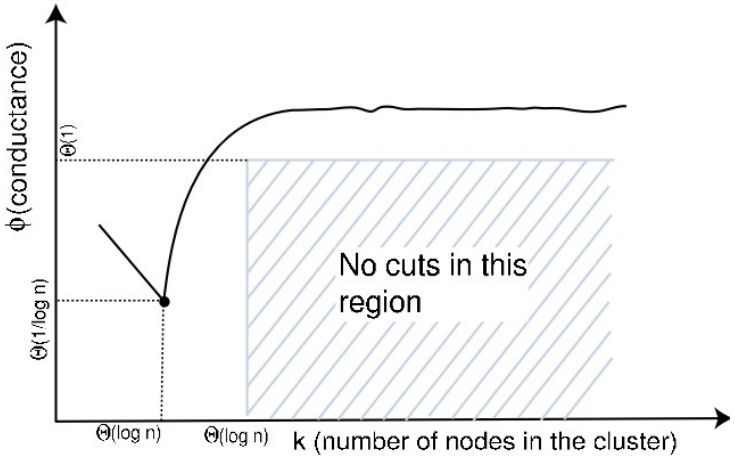


US-Senate

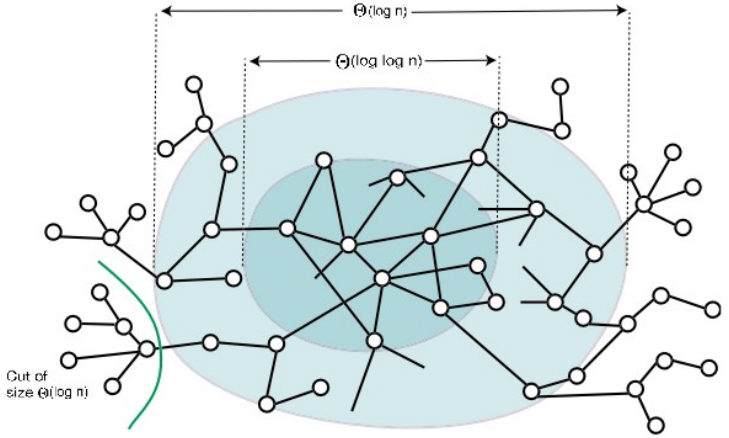
Interpretation (of upward-sloping NCP): A simple theorem on random graphs

Leskovec, Lang, Dasgupta, and Mahoney (2010)

Let $\mathbf{w} = (w_1, \ldots, w_n)$, where
$w_i = ci^{-1/(\beta-1)}, \quad \beta \in (2,3)$.
Connect nodes i and j w.p.
$p_{ij} = w_i w_j / \sum_k w_k$.



Power-law random graph with β ε (2,3).



Structure of the G(w) model, with β ε (2,3).

- **Sparsity** (coupled with randomness) **is the issue**, *not* heavy-tails.
- (Power laws with β ε (2,3) give us the appropriate sparsity.)

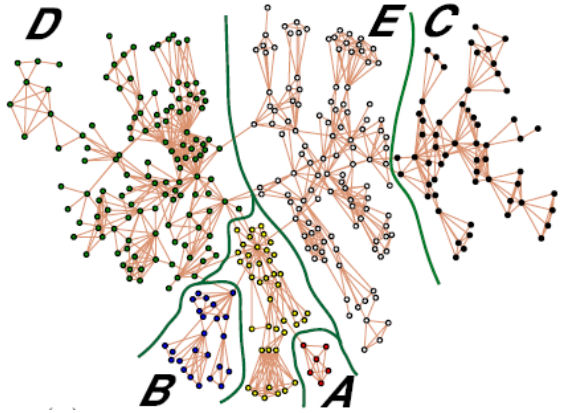
- This lack of concentration seen in NCP, cut-based, & metric-based tree-like measures.
- Data "looks like" local-structure on global-noise, not small noise on global structure

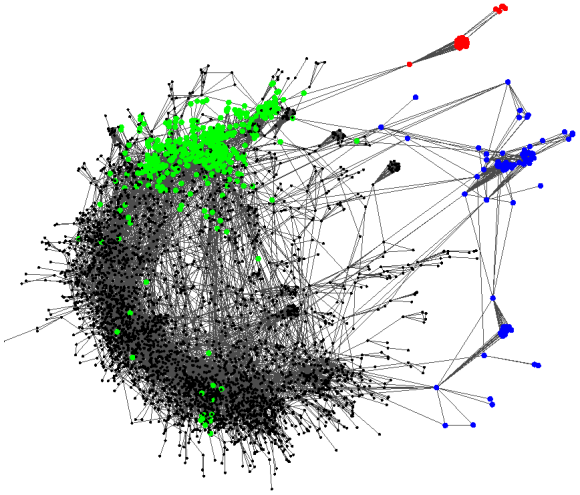
α	β
β	γ

Small versus Large Networks

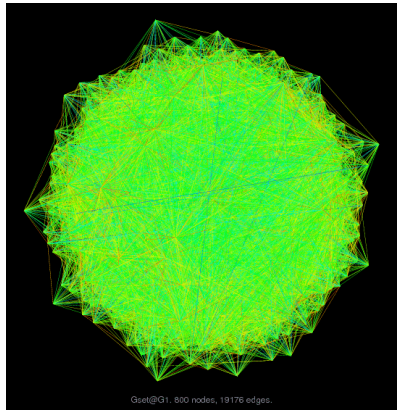
Leskovec, et al. (arXiv 2009); Mahdian-Xu 2007

- Small and large networks are very different:





(also, an expander)



E.g., fit these networks to Stochastic Kronecker Graph with "base" K=[a b; b c]:

$K_1 =$

0.99	0.17
0.17	0.82

0.99	0.55
0.55	0.15

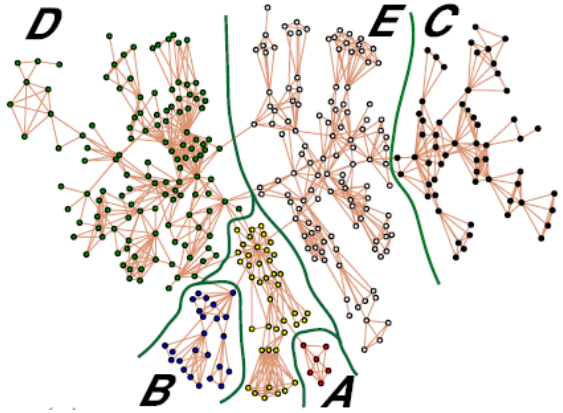
0.2	0.2
0.2	0.2

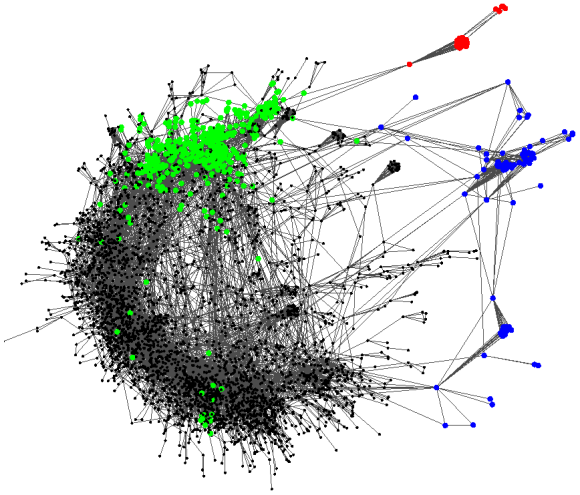
α	β
β	γ

Small versus Large Networks

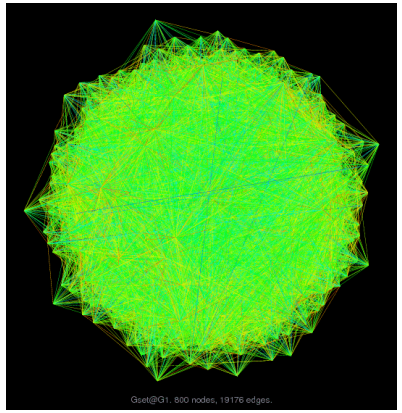
Leskovec, et al. (arXiv 2009); Mahdian-Xu 2007

- Small and large networks are very different:

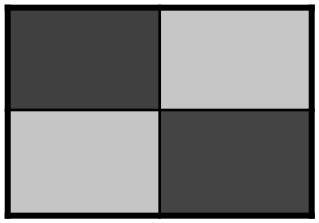


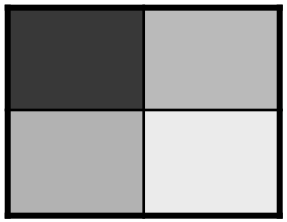


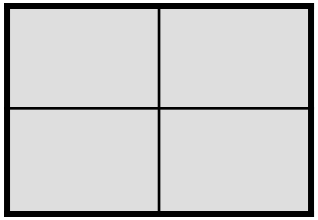
(also, an expander)



E.g., fit these networks to Stochastic Kronecker Graph with "base" K=[a b; b c]:

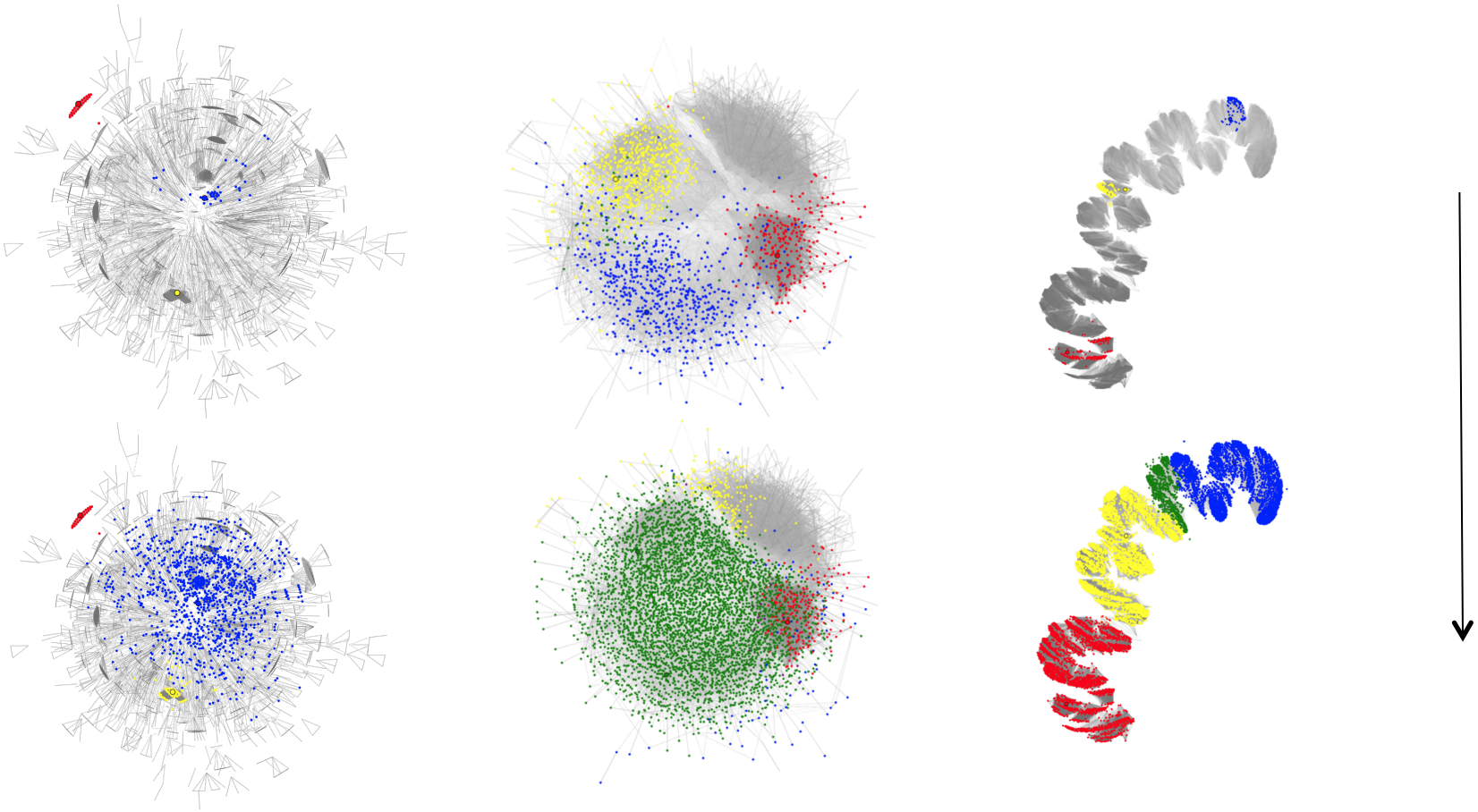
$K_1 =$ 





Small-scale to large-scale structure

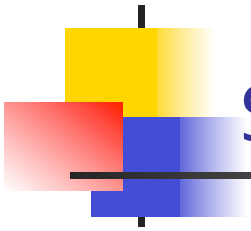
Jeub, Balachandran, Porter, Mucha, and Mahoney (2014)



CA-GrQc

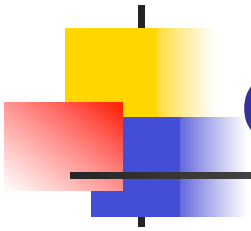
FB-Johns55

US-Senate



Summary of comparison with previous work

- Existing TD heuristics and δ hyperbolibity capture similar core-periphery structure as fast k-core hueristic
- On graphs with upward-sloping NCPs ("extremely sparse" graphs):
  - this is captured by NCP plots and stochastic Kronecker graphs, and effects of lack-of-concentration seen in δ, TDs, and NCPs
- On graphs with flat NCPs ("very sparse" graphs):
  - small (non-good-conductance) clusters still peripheral in TDs, but many more nodes in deep cores
- On graphs with downward-sloping NCPs (low-dimensional graphs):
  - tree-like core-periphery less meaningful, and there is enough heterogeneity to cause problems for TD heuristics
- Diffusion based methods (e.g., both viral propagation and ML algorithms) behave very differently, loosely captured by structure TD heuristics identify



Conclusions (on tree-like structure)

Empirical tree-like structure in many realistic informatics graphs

- for both metric-based and cut-based notions of tree-like-ness
- consistent with prior work on local-global isoperimetric NCP properties

Existing methods to quantify tree-like-ness don't do the job well

- typically for fairly subtle yet fundamental reasons.

Develop improved tree decomposition heuristics?

- More robust/scalable than hyperbolicity, and used in scientific computing
- Connections with statistical inference

Importance of intermediate-scale structure

- Mediates between very local/small-scale and very global/large-scale
- Local-global properties are a key determinant of behavior of algorithms