A Basis for Systematic Analysis and Generation of Network Topologies

Priya Mahadevan

UC San Diego

Dmitri Krioukov (CAIDA), Kevin Fall (Intel Research),

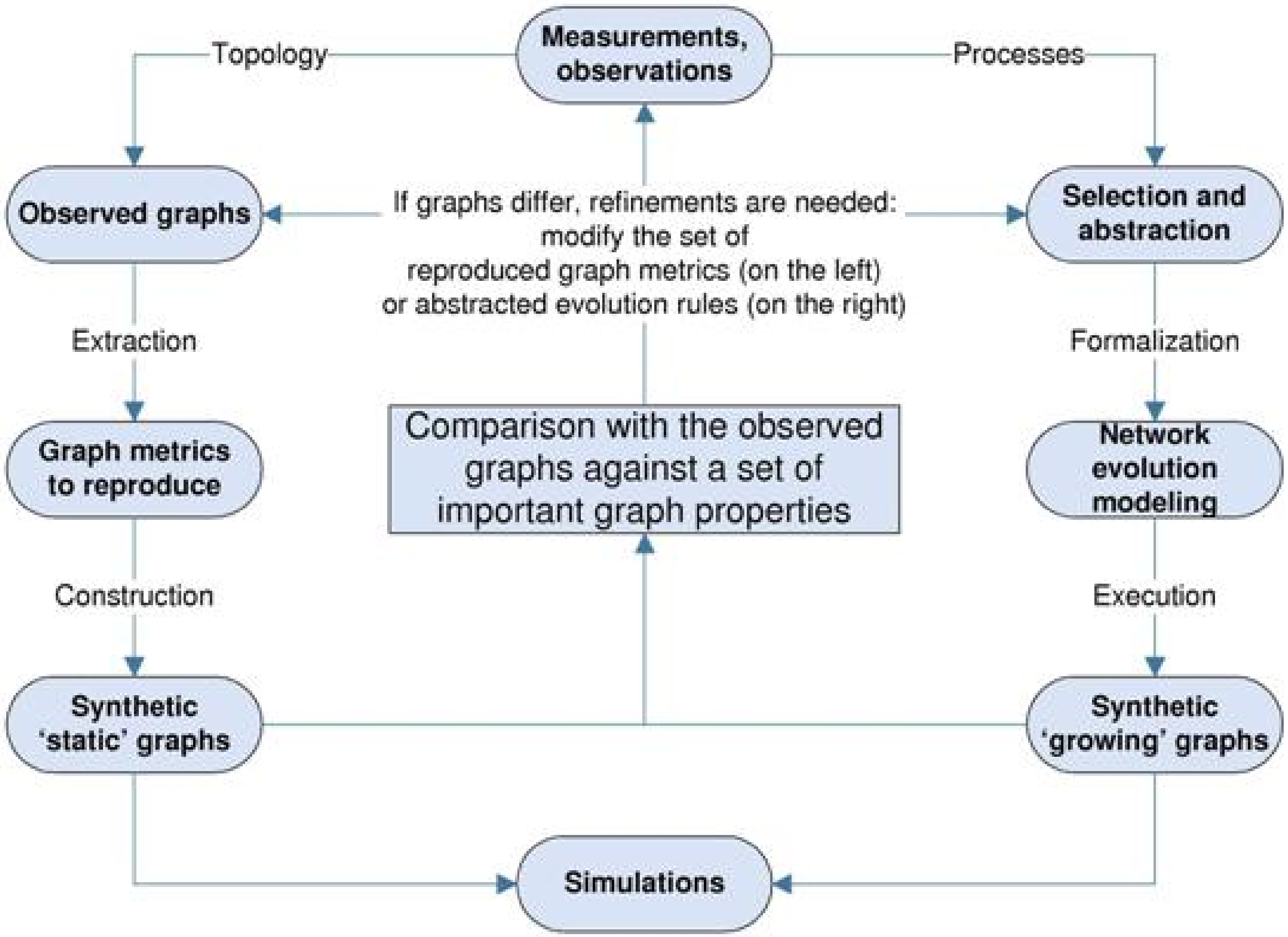
Amin Vahdat (UC San Diego)

Importance of Network Topology

- Performance of protocols and applications
  - Routing, overlay networks

    For example: new routing protocol might offer *X*-time smaller routing tables for today but scale *Y*-time worse, with *Y* >> *X*
- Robustness of the network
- Traffic engineering
- Network management
- Spread of worms, etc.

Methodologies of Topology Research



Topology Evaluation Metrics

- Distance distribution
- Betweenness
- Clustering
- Assortativity coefficient / likelihood
- Spectrum

Problem in generating graphs?

- No known techniques to produce graphs with a given form of distance distribution, betweenness, etc.
- What if a new important metric is discovered?

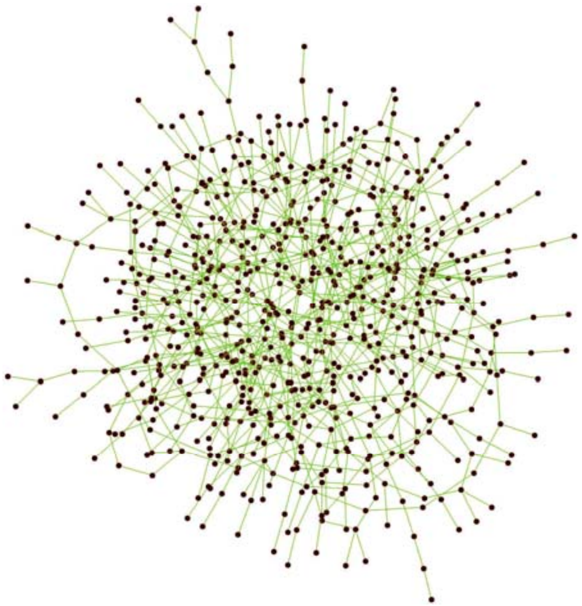
Our Approach

Enumerable set of properties P_d, $d = 0, 1, \ldots$ that satisfy:

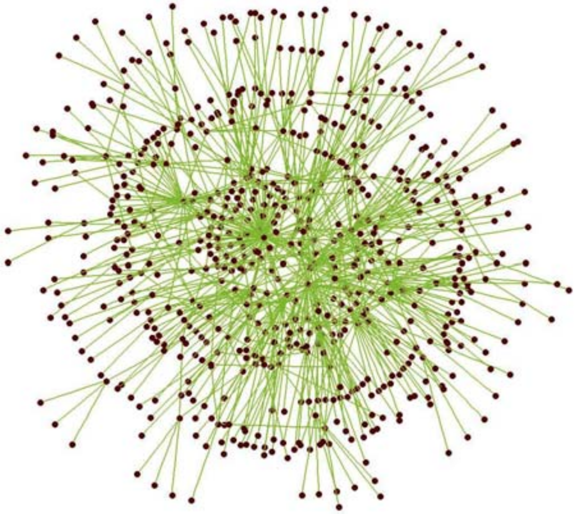
- *Constructibility*: construct graphs having these properties
- *Inclusion*: property P_d subsumes P_j where $j = 0, \ldots, d\text{-}1$
- *Convergence:* As *d* increases, the set of graphs having property P_d converges to the original graph G

Connectivity is the most basic property of network topologies

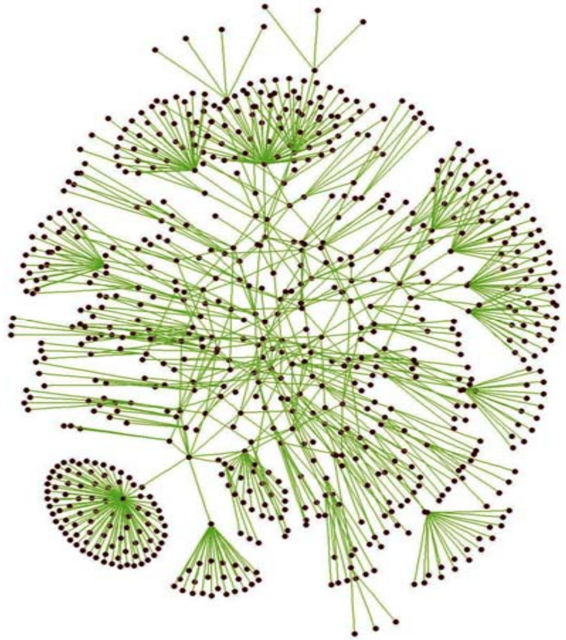
- We consider degree correlations among increasingly larger set of connected nodes
- P_0, P_1, P_2,… P_n correspond to degree correlations among connected nodes of size $0, 1, 2, \ldots n$



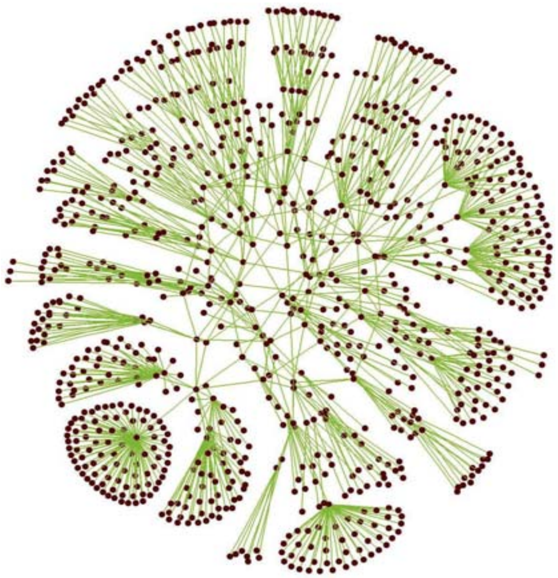
Graph reproducing P_0



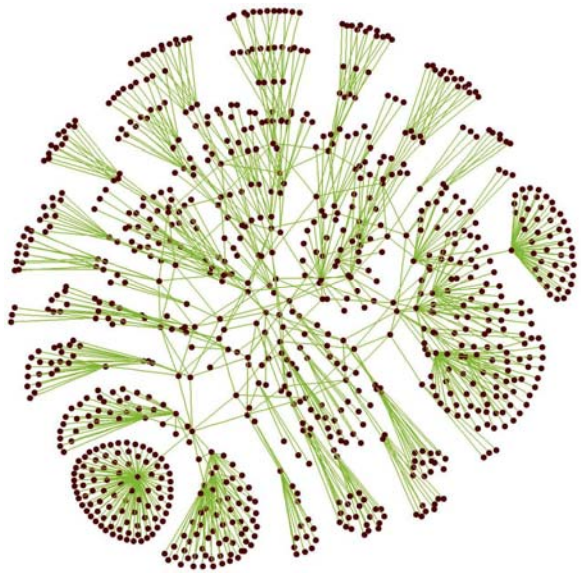
Graph reproducing P_1



Graph reproducing P_2



Graph reproducing P_3



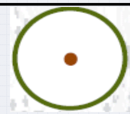
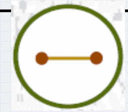
Original graph (HOT)

Outline

- Background
- Methodology
  - Graph generation
- Validation
  - AS-level graph (skitter)
  - Router-level graph (HOT)
- Limitations
- Conclusions

Our *dK*-series

dK-series: degree correlations within non-isomorphic simple connected subgraphs of size *d*

Tag	Name	Subgraphs of size:	Symbolics
0K	Average node degree	0	
1K	Node degree distribution	1	
2K	Joint degree distribution	2	
3K	Joint edge degree distribution	3	
…	…	…	…
nK	Full degree distribution	n = # of nodes	

More on *dK*-series

dK-series : degree correlations within non-isomorphic simple connected subgraphs of size *d*

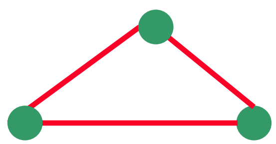
How groups of *d*-nodes with different degrees interconnect

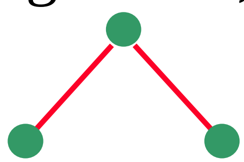
- *0K* (average degree)
- *1K* (degree distribution)
- *2K*

Connectivity between nodes having degrees *k1* and *k2*

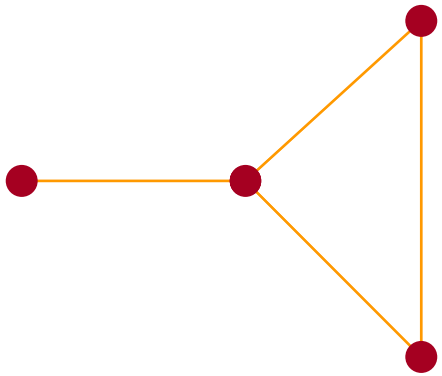
- *3K*

Connectivity among nodes having degrees *k1*, *k2* and *k3*





Example

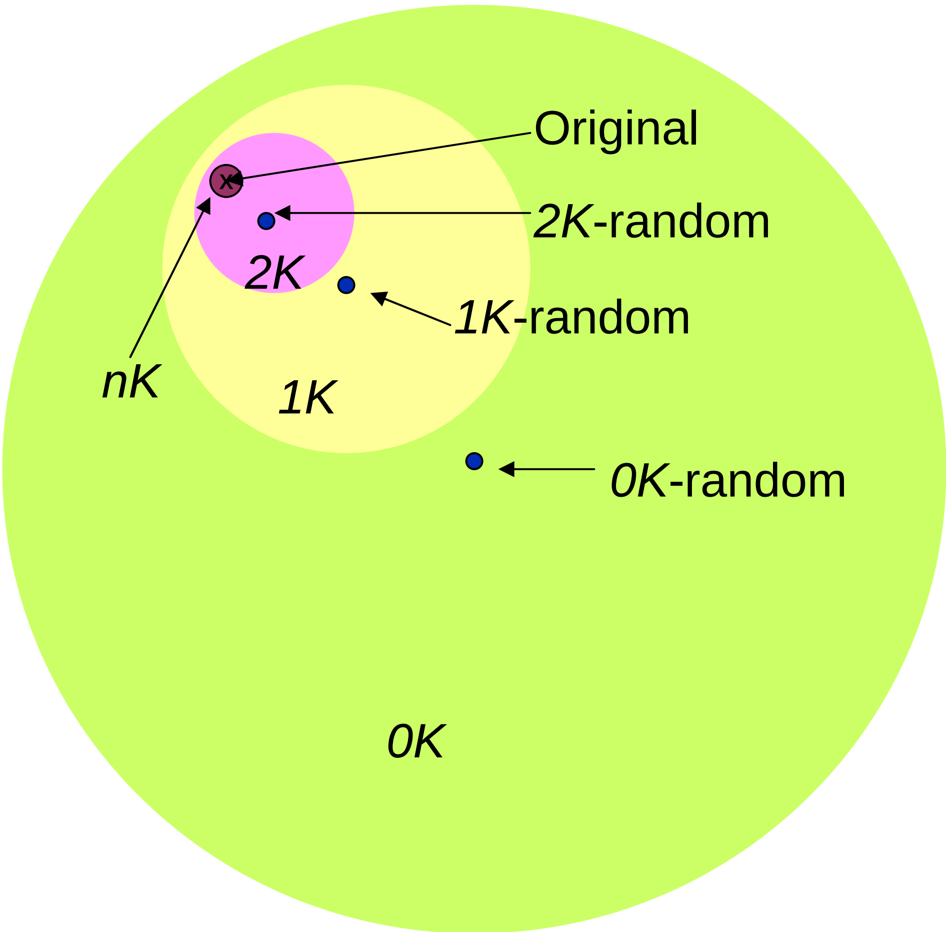


- *0K*: *avg deg* = 2
- *1K*: N(1)=1, N(2)=2, N(3)=1
- *2K*: N(1,3)=1
  N(2,2)=1
  N(2,3)=2
- *3K*: $N_{\triangle}(2,2,3)=1$
  $N_{\wedge}(1,3,2)=2$

Compute probability distributions for each

dK-graphs

- Graphs associated with each P_d
Original graph belongs to all subsets of *dK*-graphs ($d=0,..n$)
- As *d* increases, *dK* graphs converge to the original graph
- *dK*-random graph reproduces the specified property P_d, random with respect to all other properties
- What value of *d* is sufficient for practical purposes to reproduce Internet topologies?



Constructing *dK*-random graphs

- Stochastic approach
  - In theory, can be generalized to any *d;* does not work well in practice
- Pseudograph (eg: PLRG)
  - In its original form, its only for 1K
  - Extended it for *2K*, but not beyond
- Rewiring
  - Perturb the given graph (swap pairs of edges) such that property P_d is preserved

- Targeting Rewiring (Metropolis Dynamics)
  - Rewire graph such that it moves from reproducing P_{d-1} to P_d

Outline

- Background
- Methodology
  - Graph generation
- Validation
  - AS-level graph (skitter)
  - Router-level graph (HOT)
- Limitations
- Conclusions

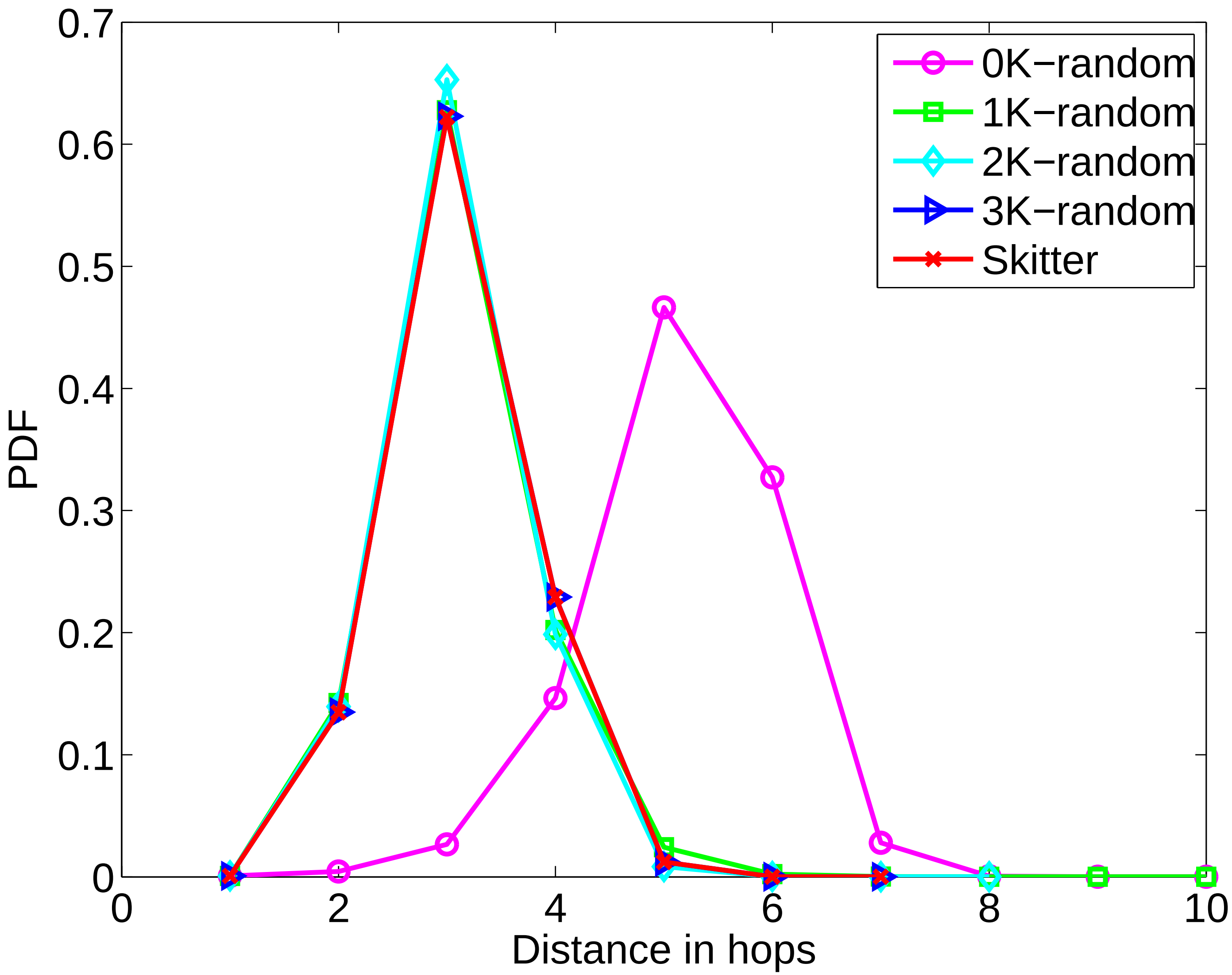
Validation

- We generated *0K*, *1K*, *2K* and *3K*-random graphs
- Compare our generated *dK*-random graphs with the original graph w.r.t. important topology metrics
- AS-level validation: skitter, BGP tables, WHOIS
- Router-level validation: HOT

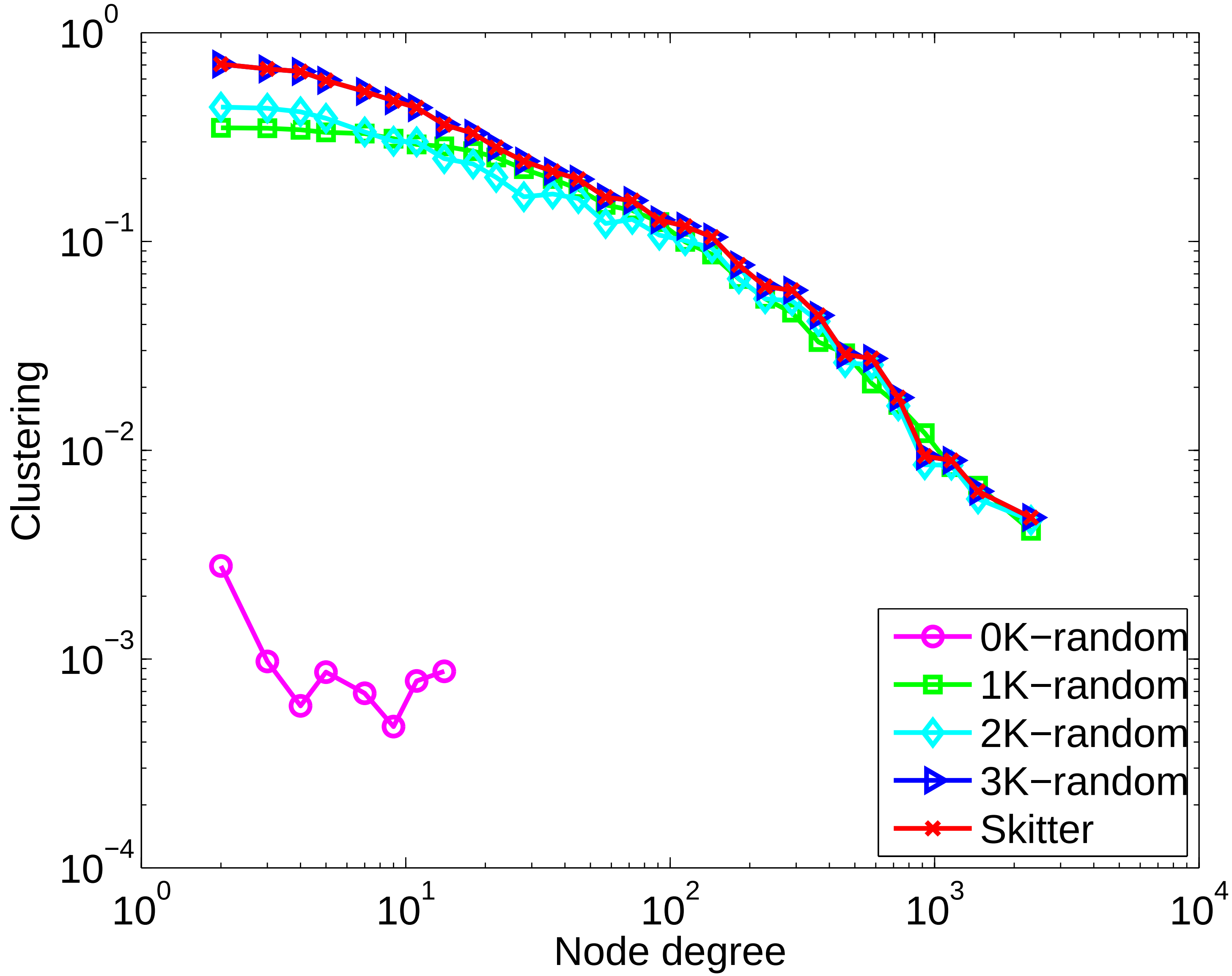
Scalar metrics comparison for skitter

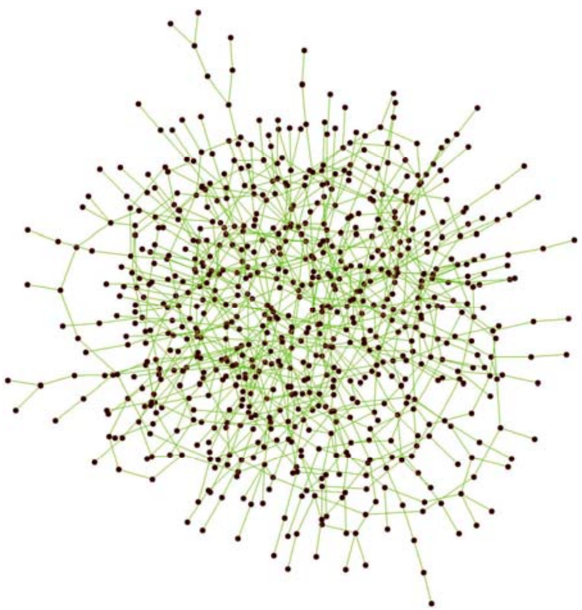
Metric	*0K*	*1K*	*2K*	*3K*	*skitter*
$<k>$	6.31	6.34	6.29	6.29	6.29
r	0	-0.24	-0.24	-0.24	-0.24
$<C>$	0.001	0.25	0.29	0.46	0.46
d	5.17	3.11	3.08	3.09	3.12
σ_d	0.27	0.4	0.35	0.35	0.37
λ_1	0.2	0.03	0.15	0.1	0.1
λ_{n-1}	1.8	1.97	1.85	1.9	1.9

Distance Distribution in skitter

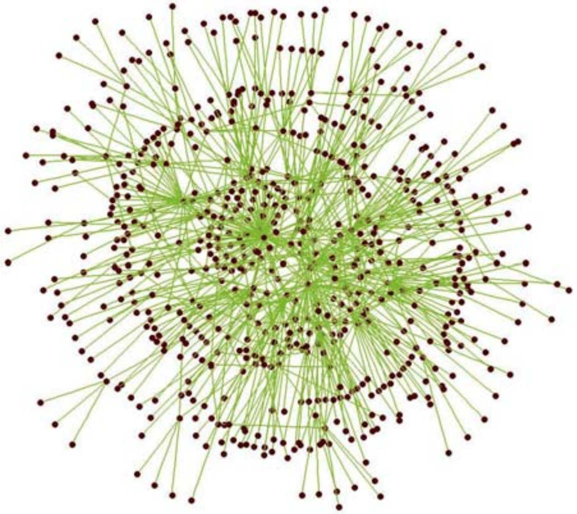


Clustering in skitter

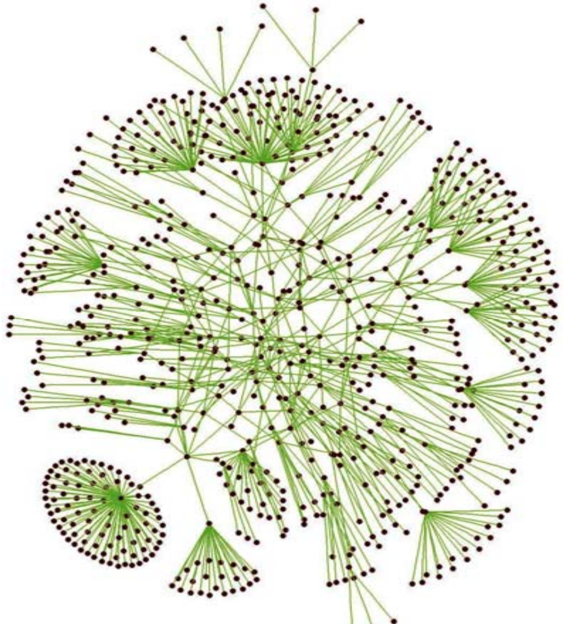




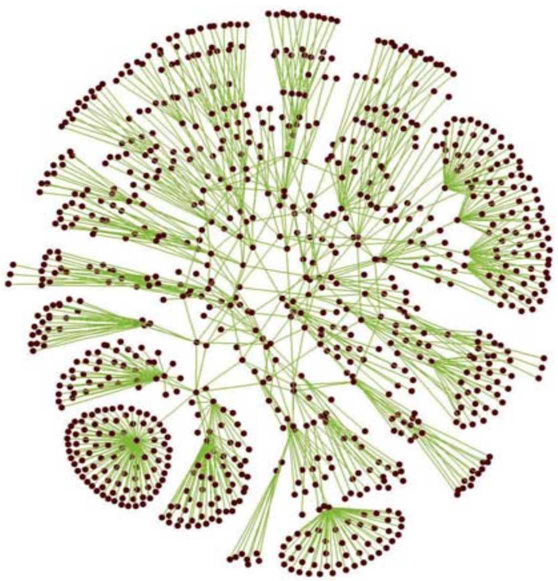
0K-random graph



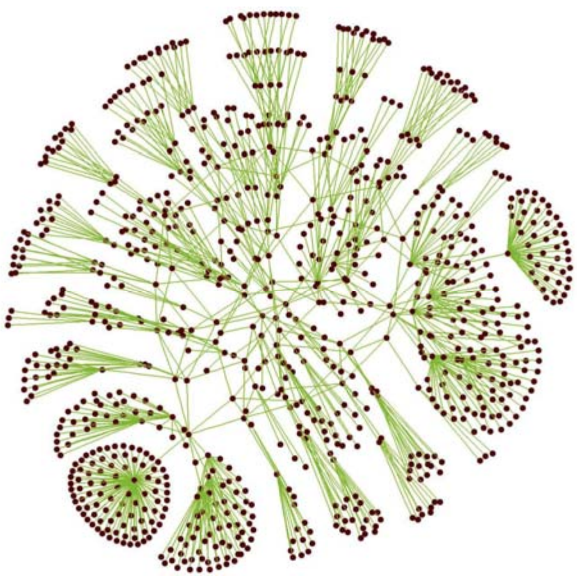
1K-random graph



2K-random graph



3K-random graph

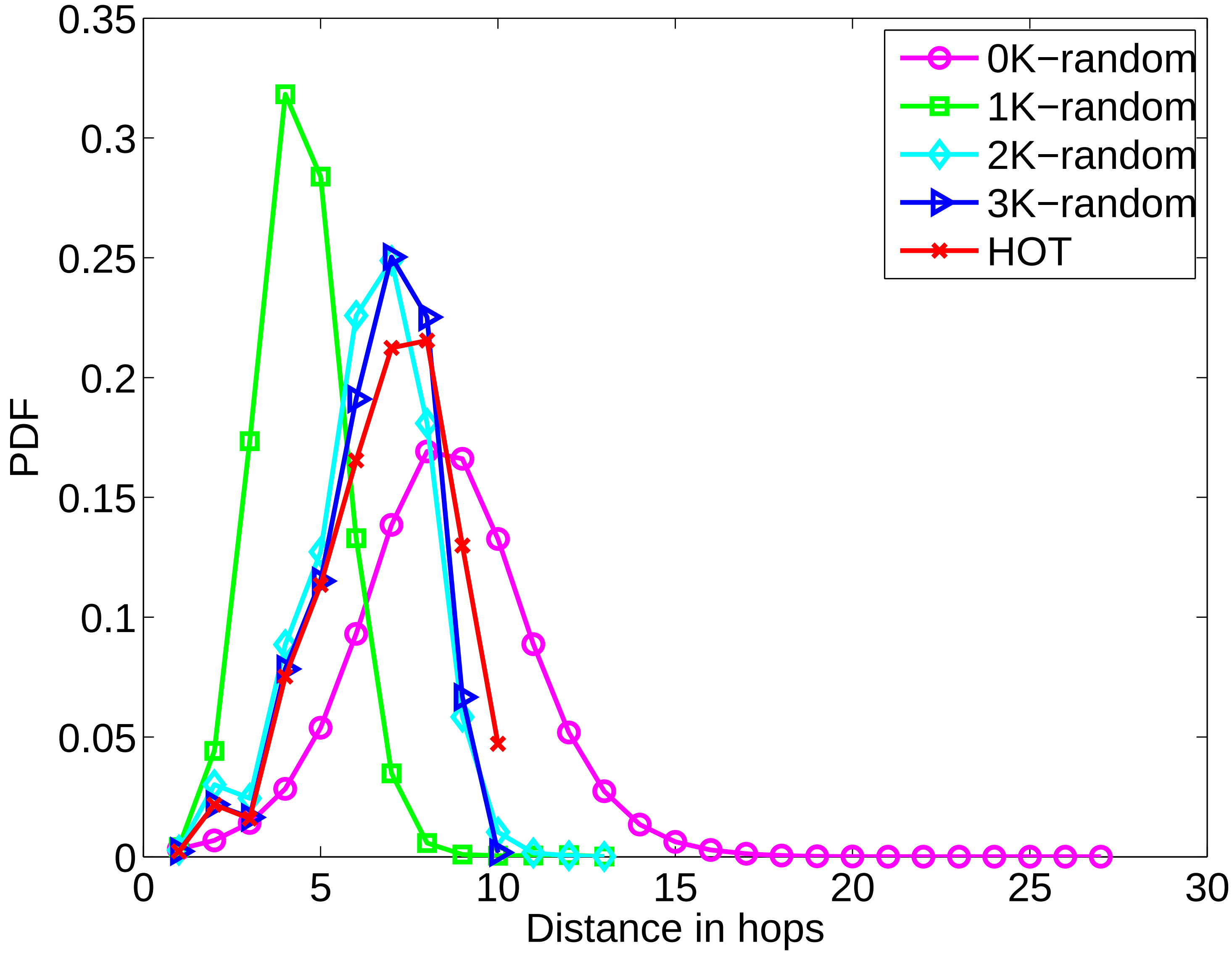


Original graph (HOT)

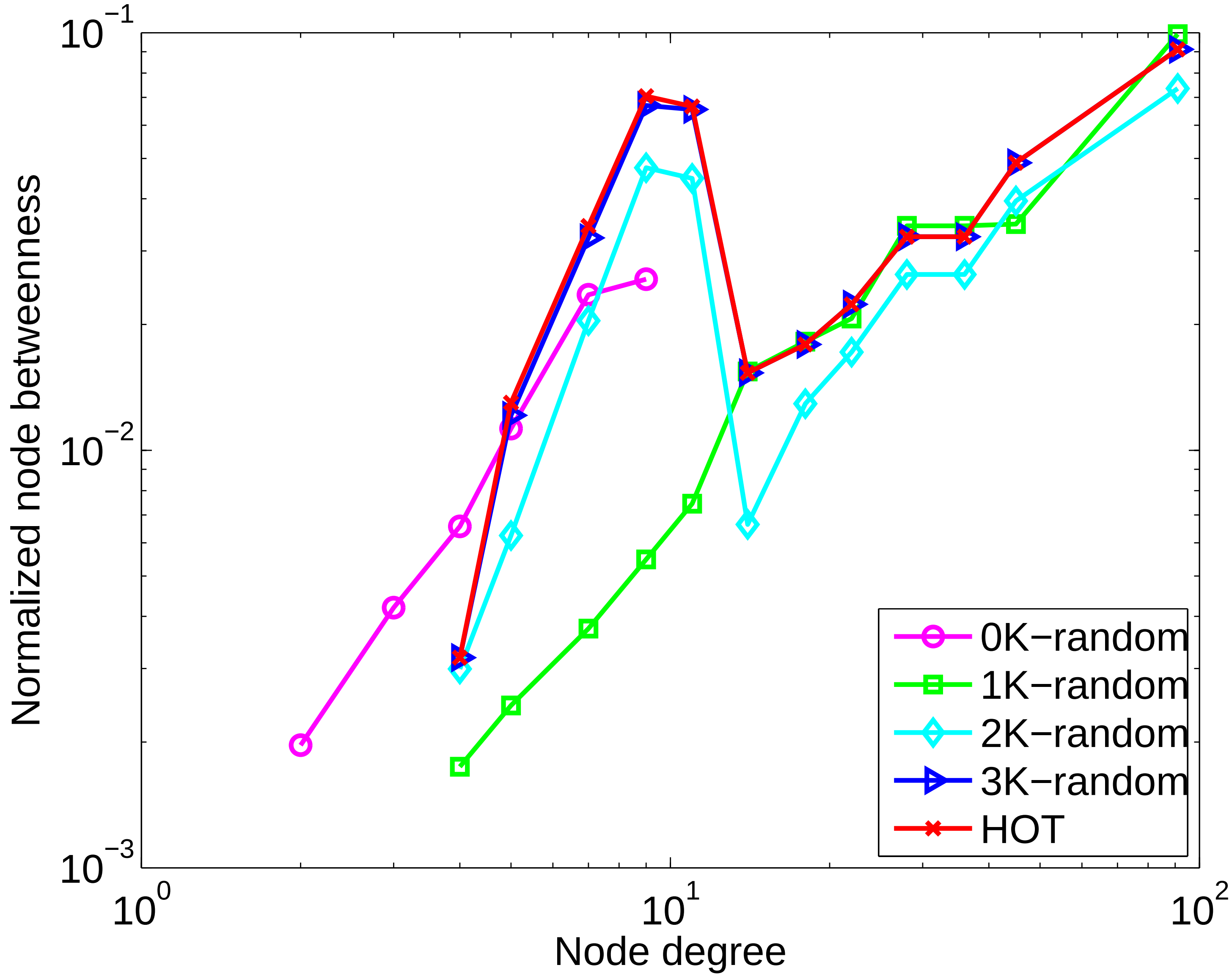
Scalar Metrics for HOT

Metric	*0K*	*1K*	*2K*	*3K*	*HOT*
$<k>$	2.47	2.59	2.18	2.10	2.10
r	-0.05	-0.14	-0.23	-0.22	-0.22
$<C>$	0.002	0.009	0.001	0	0
d	8.48	4.41	6.32	6.55	6.81
σ_d	1.23	0.72	0.71	0.84	0.57
λ_1	0.01	0.034	0.005	0.004	0.004
λ_{n-1}	1.989	1.967	1.996	1.997	1.997

Distance Distribution in HOT



Betweenness in HOT



Results summary

- For AS-level topologies:
  - *dK*-series convergence is fast
  - *2K*-random graphs capture most metrics; need *3K* for clustering
- For router-level topologies:
  - *dK*-series convergence is slower
  - *3K*-random graphs reproduce most metrics
- Router-level topologies are less random; reflect careful design and planning

Discussions and Limitations

- For graphs that we considered, $d=3$ seems sufficient
- Not all graphs can be approximated using *3K*
- What if we have to reproduce a new metric?
  - Increase d
  - Extreme case of $d = n$, generated graphs must be isomorphic to the given graph
  - Computational complexity grows rapidly with d
- Cannot discover evolutionary growth of a network

Conclusions

- No need to capture individual metrics
- Our *dK*-series , $d = 0, 1, \ldots n$ specify degree correlations within non-isomorphic simple connected subgraphs of size *d*
- By increasing *d*, we capture more complex properties of a given graph
- $d=3$ is sufficient to reproduce important metrics of observed Internet graphs

More Information

1. A Basis for Systematic Analysis of Network Topologies, *Priya Mahadevan , Dmitri Krioukov, Kevin Fall, and Amin Vahdat. ACM SIGCOMM 2006.*