On the Diversity of Graphs with High Variable Node Degrees

Lun Li
David Alderson
John C. Doyle
Walter Willinger

2006 WIT

Main Ideas of This Talk

- In general, there exist multiple graphs having the same aggregate statistics (we know this!)
- BUT, how to characterize the “diversity” among these graphs?
  - Use degree distribution as an example
  - Similar questions arise for other metrics
- Some graph theoretic metrics implicitly measure against a “background set”.
  - the nature of this background set can have serious implications for its interpretation
  - OR… are all graph theoretic measures comparable?

Some notation

Let d_i denote the degree of node i

Call $D = \{d_1, d_2, \ldots, d_n\}$ degree sequence of graph

Assume D is always ordered such that $d_1 \geq d_2 \geq \ldots \geq d_n$

We will focus on diversity among graphs having the SAME degree sequence D…

… particularly when D is scaling.

Deterministic form of scaling Relationship:

$$kd_k^{\alpha} \approx c \text{ where } 0 < c \text{ and } 0 < \alpha.$$

$$\log(k) + \alpha \log(d_k) \approx log(c)$$

Scaling and high variability

- For a sequence D,

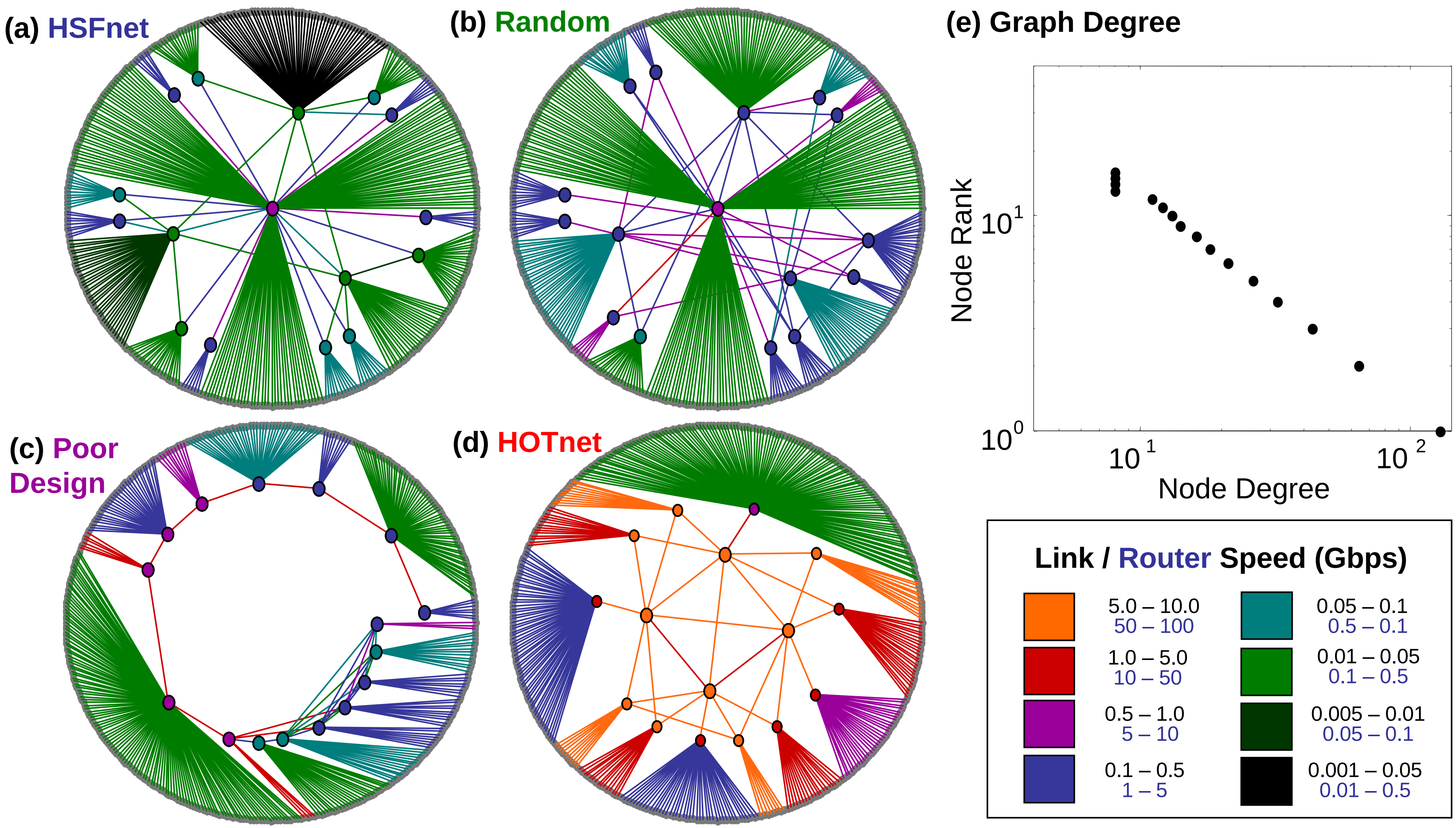
average degree: $\bar{d} = n^{-1} \sum_{k=1}^{n} d_k$

standard deviation: $STD(D) = (\sum_{k=1}^{n} (d_k - \bar{d})^2/(n-1))^{1/2}$

coefficient of variation: $CV(D) = \frac{STD(D)}{\bar{d}}$

  - If *D* is Scaling ($n \to \infty$), $\alpha < 2$, *CV(D)* = ∞
  - Star ($n \to \infty$), *CV(D)* = ∞
  - Chain ($n \to \infty$), *CV(D)* = *0*
  - If D has exponential form, *CV(D)* = Constant

Variability in the space of graphs G(D)



A Structural Approach

- s-metric

$$s(g) = \sum_{(i,j)\in\varepsilon} d_i d_j$$

- Properties:
  - Differentiate graphs with the same degree sequence
  - Depends only on the connectivity of a given graph not on the generation mechanism
  - High *s(g)* is achieved by connecting high degree nodes to each other
  - Quantify the role of the highly connected hubs

For any degree sequence D, one can construct an s_{max} Graph

- The s_{max} graph is the graph having the largest s(g)-value
- Its value depends on the “Background Set” of graphs

Impact of Background Sets on the s_{max} Graph

- Let $\mathcal{G}(D)$ denote the space of graphs having degree sequence D (could be disconnected or non-simple). One can show that within $\mathcal{G}(D)$:

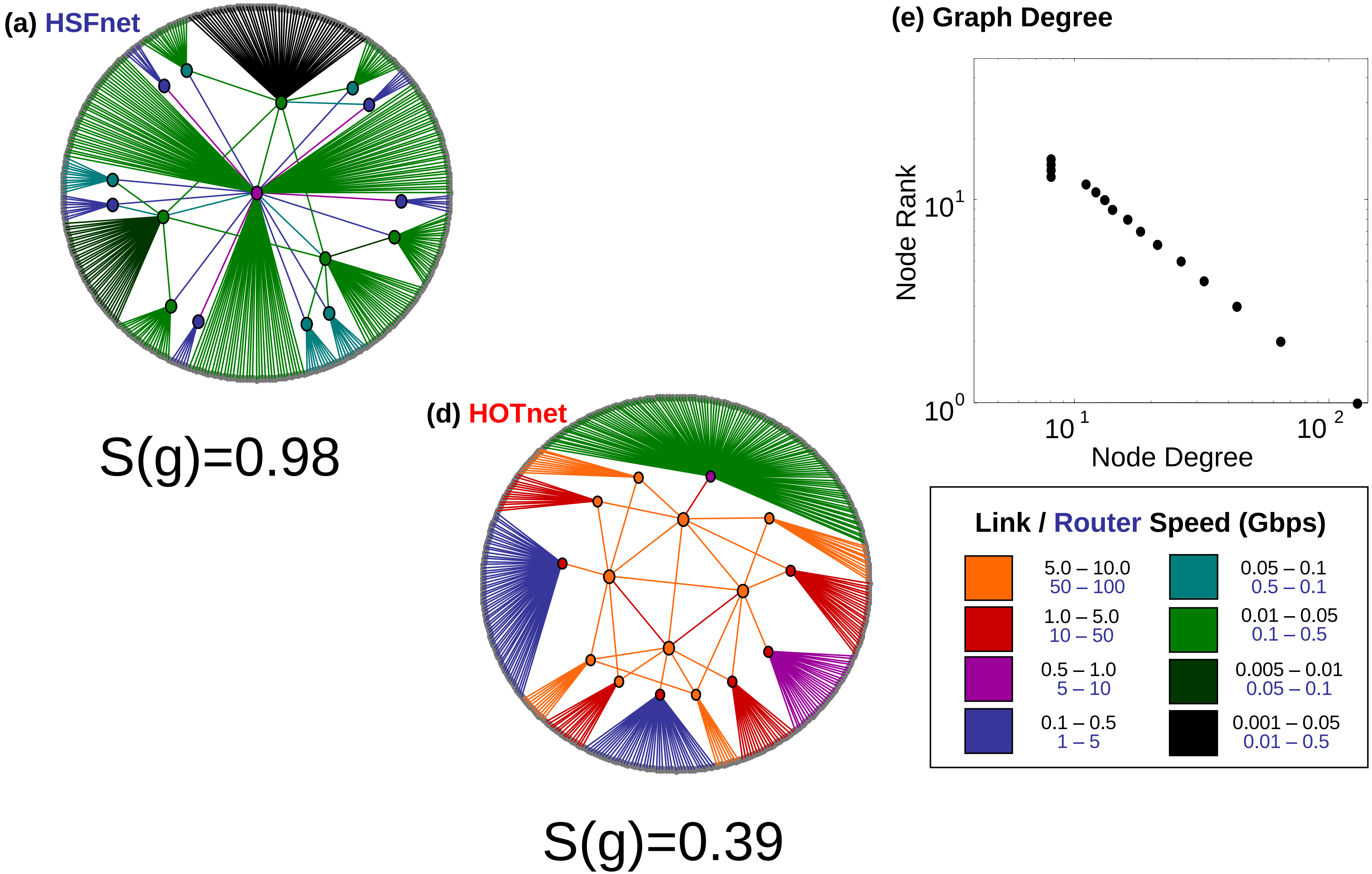
$$s_{\text{max}} = \sum_{i=1}^{n} (d_i/2) \cdot d_i^2 = \sum_{i=1}^{n} d_i^3/2$$

- Let $G(D)$ denote the space of *simple* and *connected* graphs having degree sequence D.

Among graphs in G(D) (simple, connected)

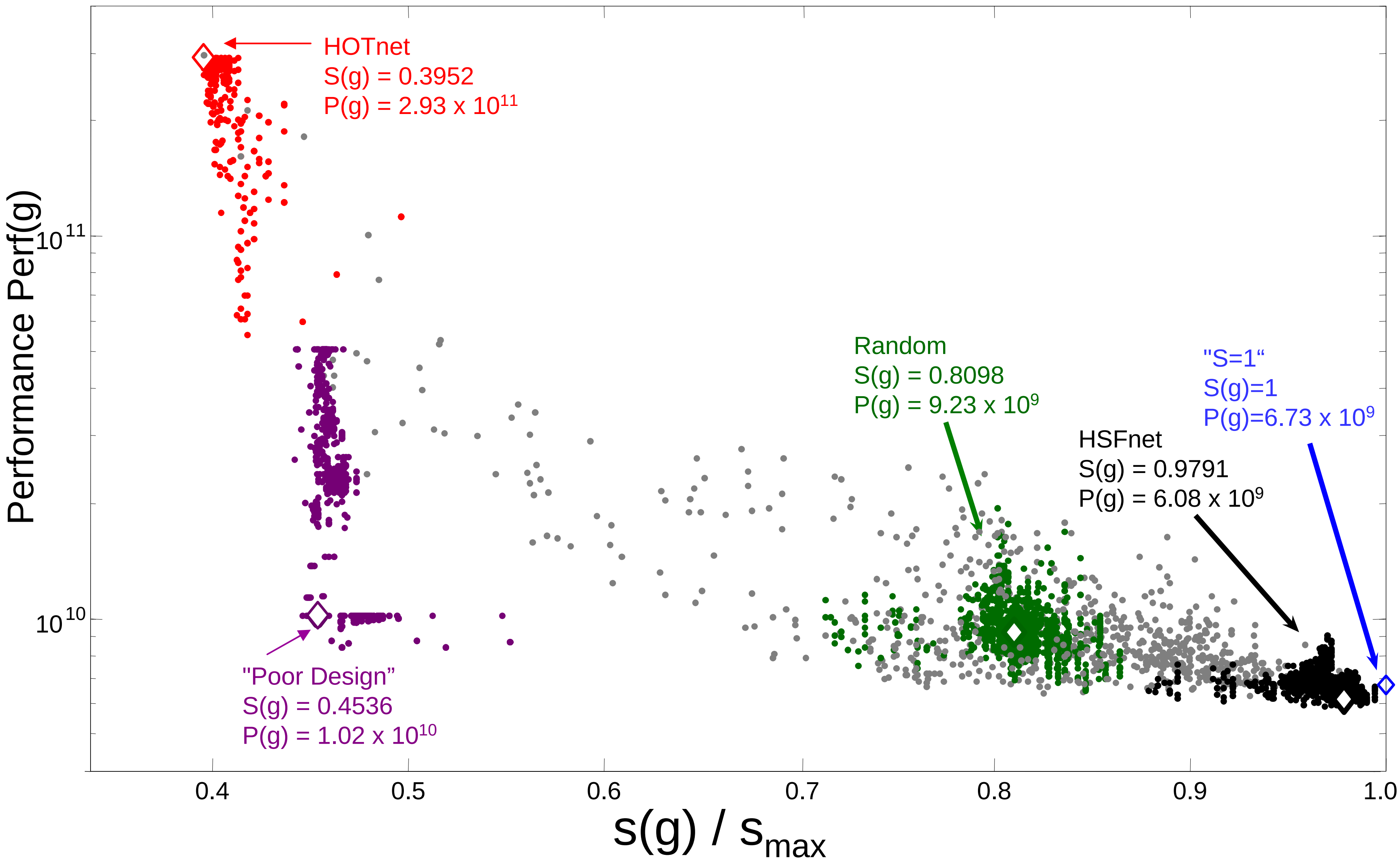
- Deterministic way to generate
- Order all potential links *(i, j)* according to their weight
- Among Acyclic graphs (trees)
  - From high degree node to low degree node

$\Rightarrow$ We will use the normalized metric: $S(g) = s(g) / s_{max}$



(a) HSFnet
S(g)=0.98
(d) HOTnet
S(g)=0.39
(e) Graph Degree
Node Rank
Node Degree
10^0
10^1
10^1
10^2
Link / Router Speed (Gbps)
5.0 – 10.0
50 – 100
1.0 – 5.0
10 – 50
0.5 – 1.0
5 – 10
0.1 – 0.5
1 – 5
0.05 – 0.1
0.5 – 0.1
0.01 – 0.05
0.1 – 0.5
0.005 – 0.01
0.05 – 0.1
0.001 – 0.05
0.01 – 0.5

Graph diversity and Perf(g) vs. s(g)



smax and graph metrics

- Node Centrality
  - In smax graph, high degree nodes have high centrality
- Self similarity
  - smax graph remains smax by trimming, coarse graining, highest connect motif
- Graph likelihood
  - smax has highest likelihood to generate by GRG
- Conjecture:
  - smax graphs are largely unique in terms of their structure

s-metric and Degree Correlations

- Assume an underlying probabilistic graph model
- Degree correlation between two adjacent vertices *k, k'* is defined as

$$P(k,k') = \frac{1}{n^2}\left\langle \sum_{i,j=1}^{n} \delta[d_i - k] a_{ij} \delta[d_j - k'] \right\rangle$$

where

$$\delta[D_i(g) - k] = \begin{cases} 1 & \text{if node } i \text{ of graph } g \text{ has degree } k \\ 0 & \text{otherwise.} \end{cases}$$

$$a_{ij} = \begin{cases} 1 & \text{if nodes } i, j \text{ are connected} \\ 0 & \text{otherwise} \end{cases}$$

- The *s*-metric is related to the degree correlation:

$$\langle s \rangle = \frac{n^2}{2} \sum_{k,k' \in D} k k' P(k,k').$$

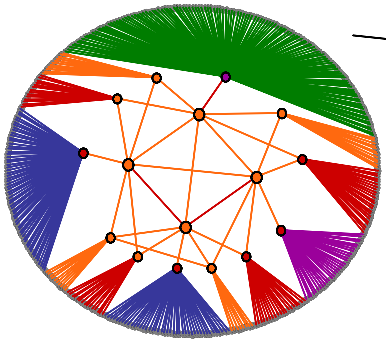
s-metric and Assortativity r(g)

- A notion of degree correlation
  - Assortative mixing: a preference for high-degree vertices to attach to other high-degree vertices
  - Disassortative mixing: the converse
- Definition [Newman]:

$$r = \frac{\sum_{k,k' \in \bar{D}} kk'(Q(k,k') - Q(k)Q(k'))}{\sum_{k,k' \in \bar{D}} kk'(Q(k)\delta[k-k'] - Q(k)Q(k'))} \qquad r \in [-1,1]$$

- r>0, assortatitive, social networks
- r<0, disassortitive, internet, biology networks

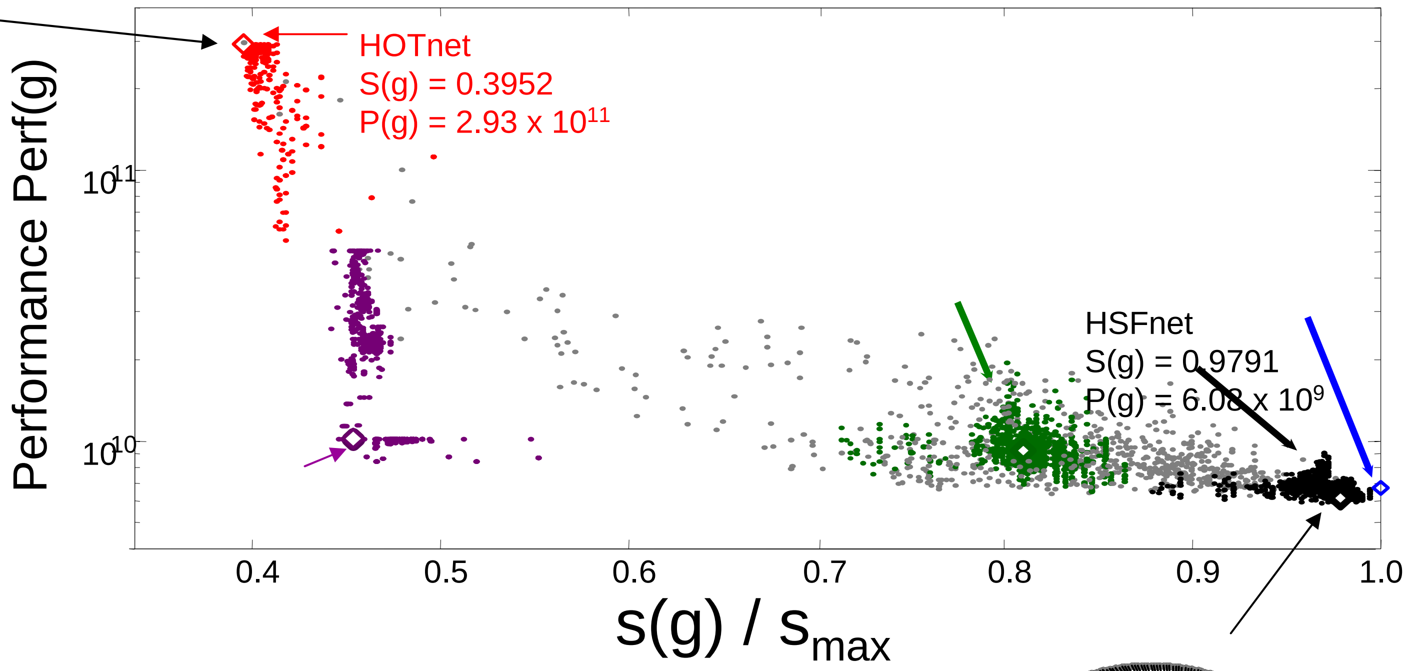
Graph diversity and r(g) vs. s(g)



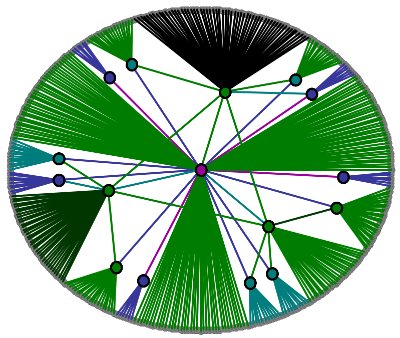
Perf = 2.93X10e11

S(g)=0.39

r = -0.4815,



Both graphs have same degree distribution and very similar assortativity!!!



Perf = 6.06X10e9

S(g)=0.98

r = -0.4283,

Assortativity *r(g)*

- For a given graph, assortativity is:

$$r(g) = \frac{\left[\sum_{(i,j)\in\mathcal{E}} d_i d_j\right] - \left[\sum_{i\in\mathcal{V}} \frac{1}{2} d_i^2\right]^2 / l}{\left[\sum_{i\in\mathcal{V}} \frac{1}{2} d_i^3\right] - \left[\sum_{i\in\mathcal{V}} \frac{1}{2} d_i^2\right]^2 / l}$$

- Normalization term

$$\left[\sum_{i\in\mathcal{V}} \frac{1}{2} d_i^3\right]$$

s_{max} of unconstrained graph

- Centering term

$$\left[\sum_{i\in\mathcal{V}} \frac{1}{2} d_i^2\right]^2 / l$$

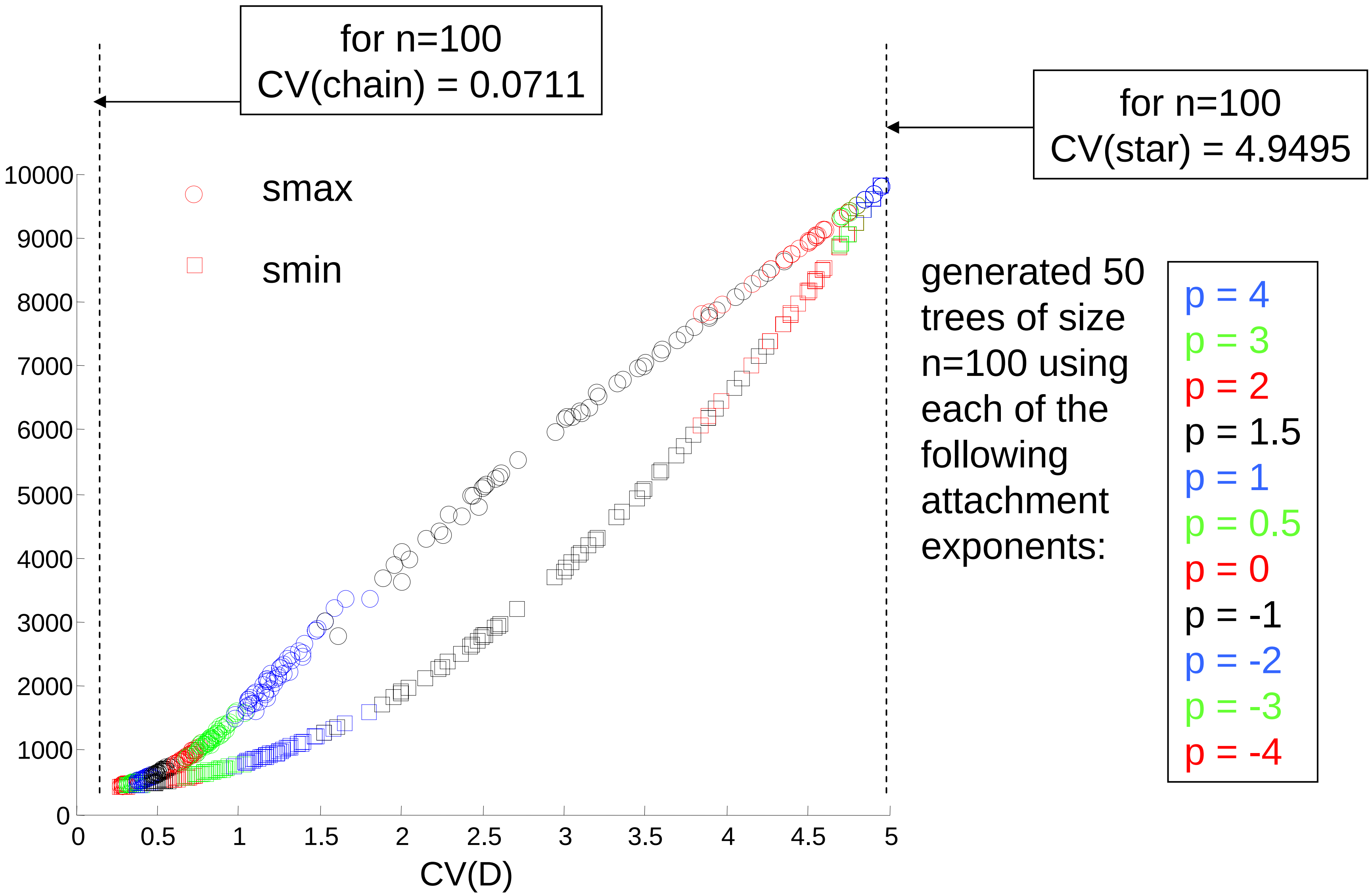
Center of unconstrained graph

- r=1: all the nodes connect to themselves
- r=-1: depends on the degree sequence
- Background set is the unconstrained graph!

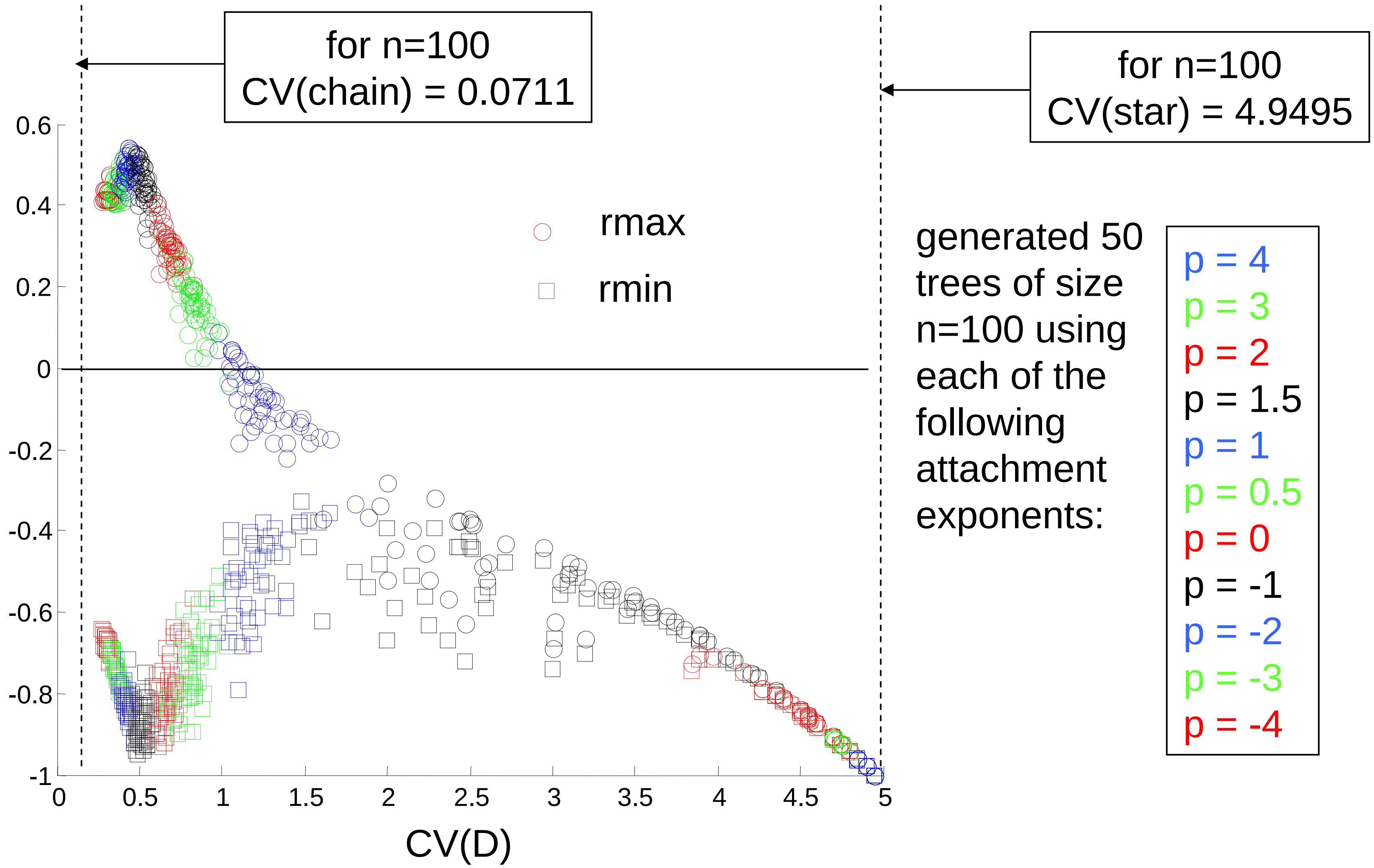
simple experiment

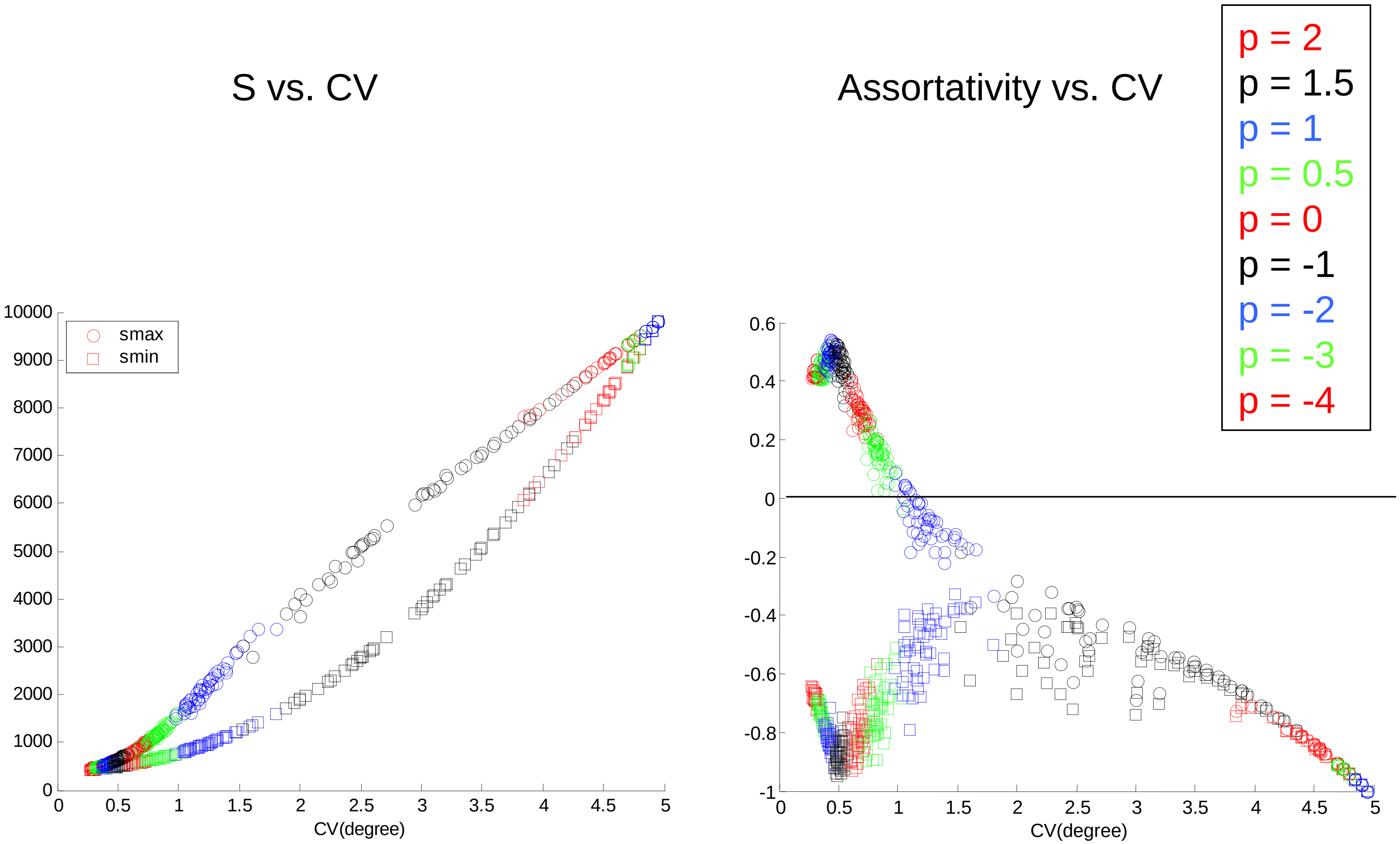
- generate multiple trees by adding a node k to an existing node j, with probability $\Pi(j) \propto (d_j)^p$
  - $p=1$ ⇔ linear preferential attachment
  - $p=0$ ⇔ uniform attachment
  - $p \to \infty$ ⇔ attach to max degree node (result = a star)
  - $p \to -\infty$ ⇔ attach to min degree node (result = a chain)
- each trial results in a tree having
  - its own degree sequence D, s-value, CV(D)
  - its own s_{min} and s_{max} values (from D),
  - its own r_{min} and r_{max} values (from D)

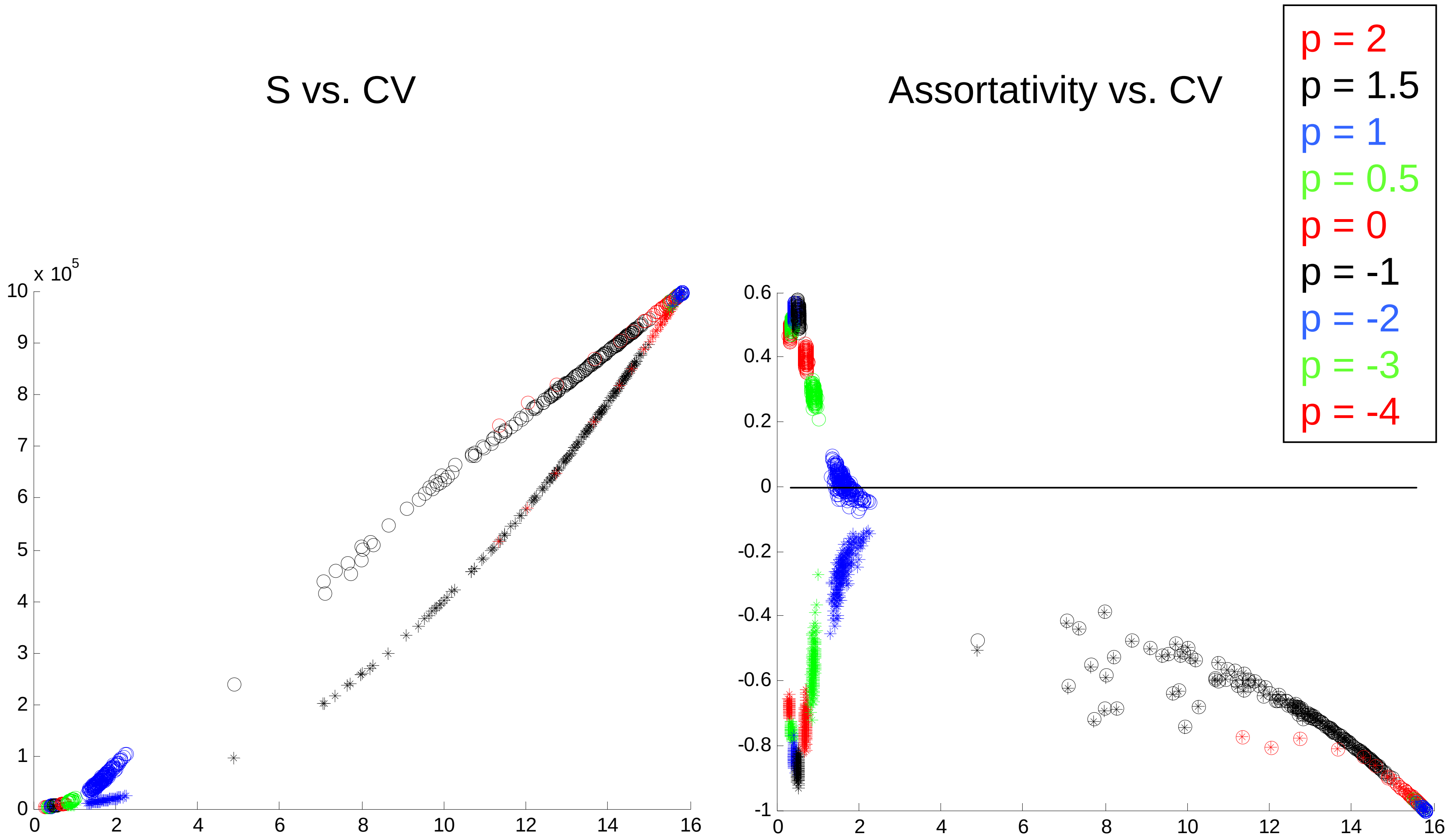
$$CV(chain) = \frac{n^{1/2}(n-2)^{1/2}}{2^{1/2}(n-1)^{3/2}} \qquad CV(star) = \frac{n^{1/2}(n-2)}{2(n-1)}$$



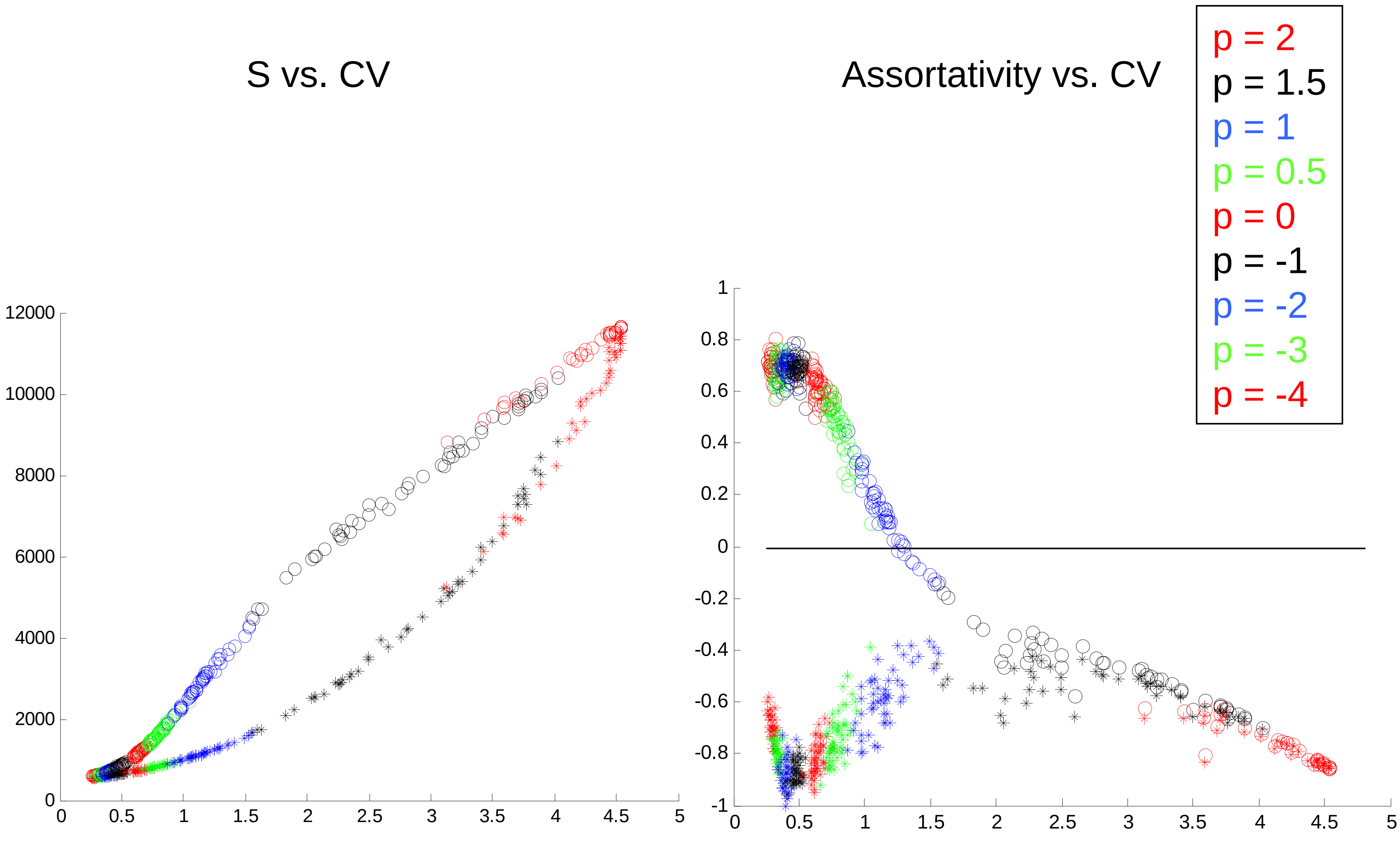
$$CV(chain) = \frac{n^{1/2}(n-2)^{1/2}}{2^{1/2}(n-1)^{3/2}} \qquad CV(star) = \frac{n^{1/2}(n-2)}{2(n-1)}$$



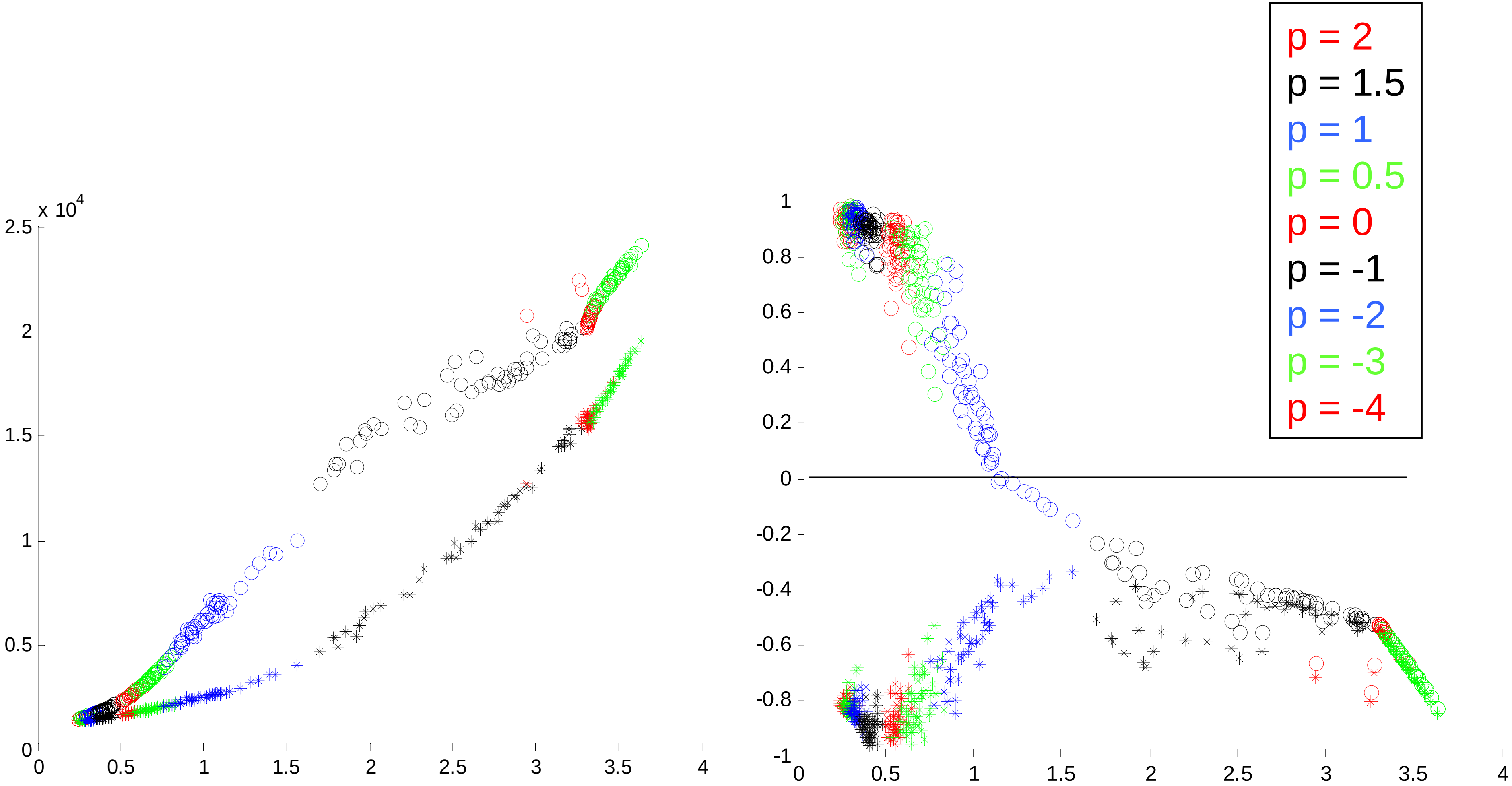

S vs. CV
smax
smin
CV(degree)
Assortativity vs. CV
CV(degree)
p = 2
p = 1.5
p = 1
p = 0.5
p = 0
p = -1
p = -2
p = -3
p = -4



S vs. CV
Assortativity vs. CV
x 10^5
p = 2
p = 1.5
p = 1
p = 0.5
p = 0
p = -1
p = -2
p = -3
p = -4



S vs. CV
Assortativity vs. CV
12000
10000
8000
6000
4000
2000
0
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
p = 2
p = 1.5
p = 1
p = 0.5
p = 0
p = -1
p = -2
p = -3
p = -4


x 10^4
2.5
2
1.5
1
0.5
0
0
0.5
1
1.5
2
2.5
3
3.5
4
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
p = 2
p = 1.5
p = 1
p = 0.5
p = 0
p = -1
p = -2
p = -3
p = -4

Conclusions

- The set *G(D)* of graphs *g* with fixed scaling degree *D* can be extremely diverse
- s-metric can highlight the difference of the graphs in *G(D).*
- *s*-metric has a rich connection to self-similarity, likelihood, betweeness and assortativity
- the nature of this background set can have serious implications for its interpretation
- These issues apply to metrics other than simple degree sequence (e.g., two-point degree correlations)

Towards a Theory of Scale-Free Graphs: Definition, Properties, and Implications
Lun Li, David Alderson, John C. Doyle, Walter Willinger
Internet Mathematics. In Press. (2006)