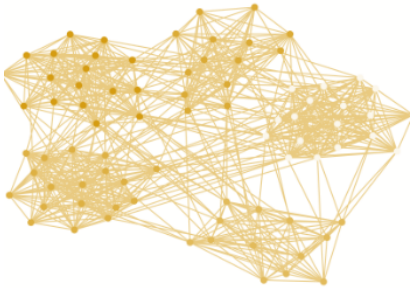
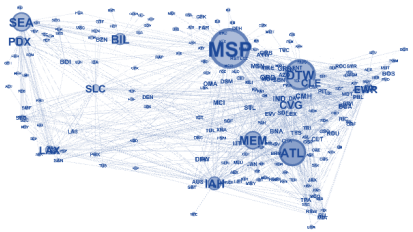
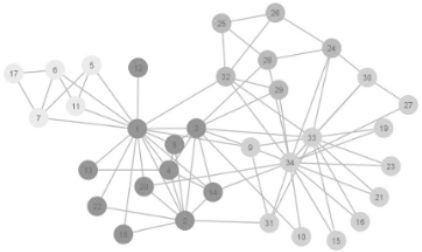
Prof. Ralucca Gera, rgera@nps.edu
Applied Mathematics Department,
Naval Postgraduate School





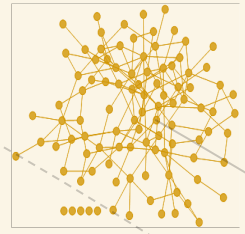


MA4404 Complex Networks

Clustering Coefficient

Learning Outcomes

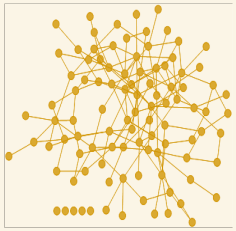
- Differentiate between:
  - Local clustering coefficient,
  - Global clustering coefficient, and
  - Average clustering coefficient
- Understand the computation of clustering coefficients
- Create extensions of clustering coefficient



$$C(G) = \frac{\#\ of\ K_3}{\#\ of\ connected\ \ triples}$$

Definition of Clustering Coefficient

Clustering coefficients for real networks



- **The clustering coefficients** measure the average probability that two neighbors of a vertex are themselves neighbors (a measure of the density of triangles in a network).
- There are three versions:
  1. Clustering coefficient of the graph (overall network clustering):

$$C(G) = \frac{\#\ of\ K_3}{\#\ of\ connected\ \ triples}$$

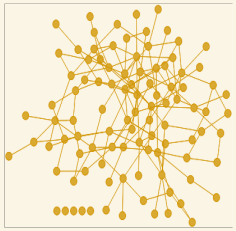
  2. Local Clustering coefficient (locally dense communities):

$$\mathrm{C_i} = \frac{\#\ of\ K_3\ that\ inlude\ i}{\#\ of\ connected\ \ triples\ centered\ at\ i} = \frac{number\ of\ edges\ between\ neighbors\ of\ i}{\deg i\ \ (\deg i\ -1)/2}$$

  3. Avg. clustering coefficient of G:

$$C_{ws}(G) = \frac{1}{n}\sum_{i\in V(G)} C_i$$

Clustering coeff distribution example in Gephi



HTML Report

Clustering Coefficient Metric Report

Parameters:

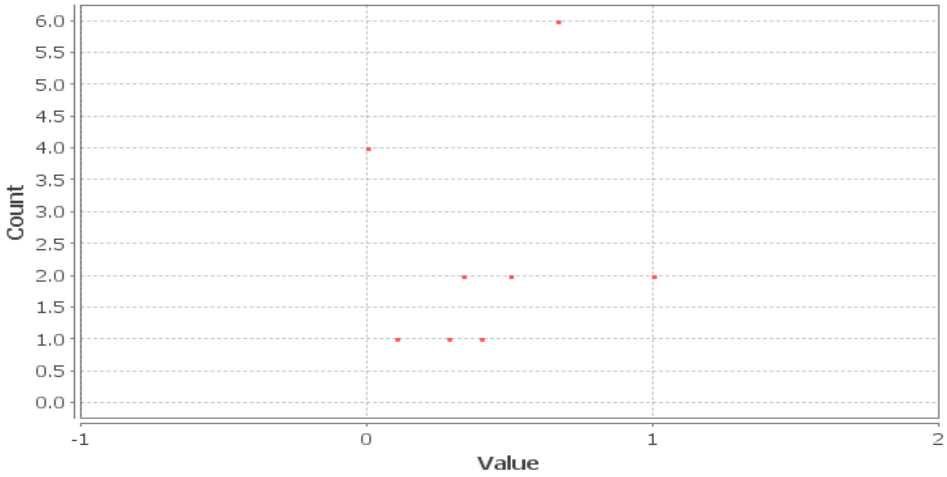
Network Interpretation: undirected

Results:

$$C_i = \frac{\#\ of\ K_3\ that\ inlude\ i}{\#\ of\ connected\ \ triples\ centered\ at\ i}$$

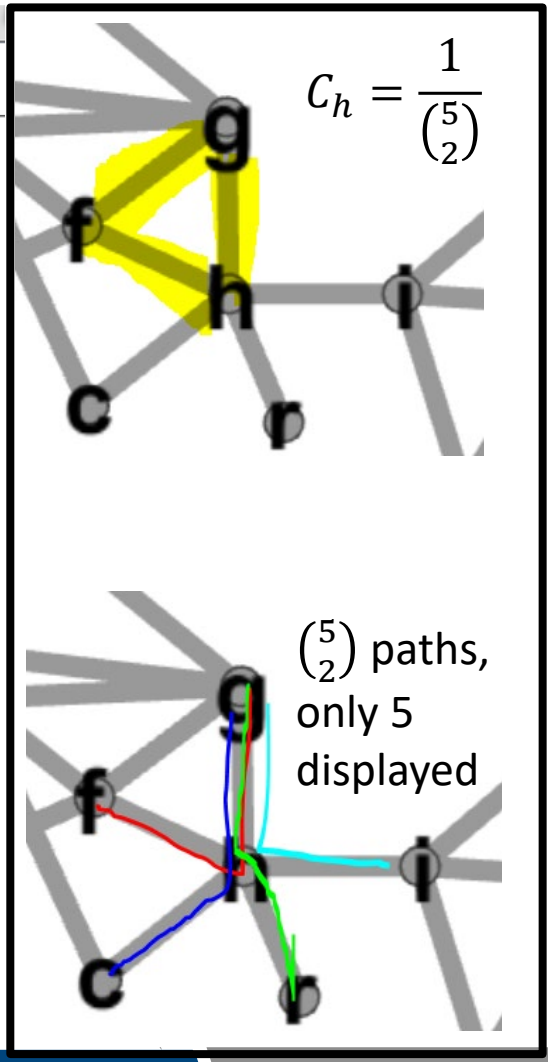
Average Clustering Coefficient: 0.497
Total triangles: 14
The Average Clustering Coefficient is the mean value of individual coefficients.

Clustering Coefficient Distribution



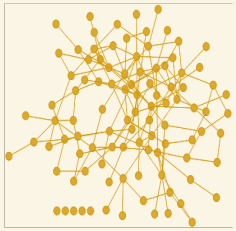
Algorithm:

Matthieu Latapy, *Main-memory Triangle Computations for Very Large (Sparse (Power-Law)) Graphs*, in Theoretical Computer Science (TCS) 407 (1-3), pages 458-473, 2008



Observed: Networks are globally sparse, but locally dense

Statistics for real networks



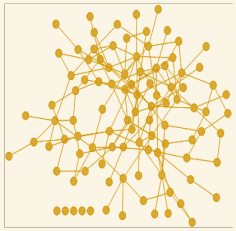
	Network	Type	n	m	c	S	ℓ	α	C	C_{WS}	
Social	Film actors	Undirected	449 913	25 516 482	113.43	0.980	3.48	2.3	0.20	0.78	0.
	Company directors	Undirected	7 673	55 392	14.44	0.876	4.60	–	0.59	0.88	0.
	Math coauthorship	Undirected	253 339	496 489	3.92	0.822	7.57	–	0.15	0.34	0.
	Physics coauthorship	Undirected	52 909	245 300	9.27	0.838	6.19	–	0.45	0.56	0.
	Biology coauthorship	Undirected	1 520 251	11 803 064	15.53	0.918	4.92	–	0.088	0.60	0.
	Telephone call graph	Undirected	47 000 000	80 000 000	3.16			2.1			
	Email messages	Directed	59 812	86 300	1.44	0.952	4.95	1.5/2.0		0.16	
	Email address books	Directed	16 881	57 029	3.38	0.590	5.22	–	0.17	0.13	0.
	Student dating	Undirected	573	477	1.66	0.503	16.01	–	0.005	0.001	−0.
	Sexual contacts	Undirected	2 810					3.2			
Information	WWW nd.edu	Directed	269 504	1 497 135	5.55	1.000	11.27	2.1/2.4	0.11	0.29	−0.
	WWW AltaVista	Directed	203 549 046	1 466 000 000	7.20	0.914	16.18	2.1/2.7			
	Citation network	Directed	783 339	6 716 198	8.57			3.0/–			
	Roget's Thesaurus	Directed	1 022	5 103	4.99	0.977	4.87	–	0.13	0.15	0.
	Word co-occurrence	Undirected	460 902	16 100 000	66.96	1.000		2.7		0.44	
Technological	Internet	Undirected	10 697	31 992	5.98	1.000	3.31	2.5	0.035	0.39	−0.
	Power grid	Undirected	4 941	6 594	2.67	1.000	18.99	–	0.10	0.080	−0.
	Train routes	Undirected	587	19 603	66.79	1.000	2.16	–		0.69	−0.
	Software packages	Directed	1 439	1 723	1.20	0.998	2.42	1.6/1.4	0.070	0.082	−0.
	Software classes	Directed	1 376	2 213	1.61	1.000	5.40	–	0.033	0.012	−0.
	Electronic circuits	Undirected	24 097	53 248	4.34	1.000	11.05	3.0	0.010	0.030	−0.
	Peer-to-peer network	Undirected	880	1 296	1.47	0.805	4.28	2.1	0.012	0.011	−0.
Biological	Metabolic network	Undirected	765	3 686	9.64	0.996	2.56	2.2	0.090	0.67	−0.
	Protein interactions	Undirected	2 115	2 240	2.12	0.689	6.80	2.4	0.072	0.071	−0.
	Marine food web	Directed	134	598	4.46	1.000	2.05	–	0.16	0.23	−0.
	Freshwater food web	Directed	92	997	10.84	1.000	1.90	–	0.20	0.087	−0.
	Neural network	Directed	307	2 359	7.68	0.967	3.97	–	0.18	0.28	−0.

C= clustering coefficient

C_{ws}= ave clustering coefficient

Newman, "The Structure and Function of Complex Networks" http://epubs.siam.org/doi/pdf/10.1137/S0036144503424480

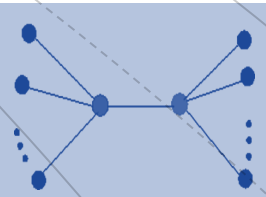
Observed vs. random



Network	n	z	C_{ws} measured	C_{ws} for random graph
Internet [153]	6,374	3.8	0.24	0.00060
World Wide Web (sites) [2]	153,127	35.2	0.11	0.00023
power grid [192]	4,941	2.7	0.080	0.00054
biology collaborations [140]	1,520,251	15.5	0.081	0.000010
mathematics collaborations [141]	253,339	3.9	0.15	0.000015
film actor collaborations [149]	449,913	113.4	0.20	0.00025
company directors [149]	7,673	14.4	0.59	0.0019
word co-occurrence [90]	460,902	70.1	0.44	0.00015
neural network [192]	282	14.0	0.28	0.049
metabolic network [69]	315	28.3	0.59	0.090
food web [138]	134	8.7	0.22	0.065

Source: N. Przulj. Graph theory analysis of protein-protein interactions. 2005.

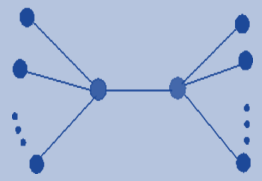
Why are they different?

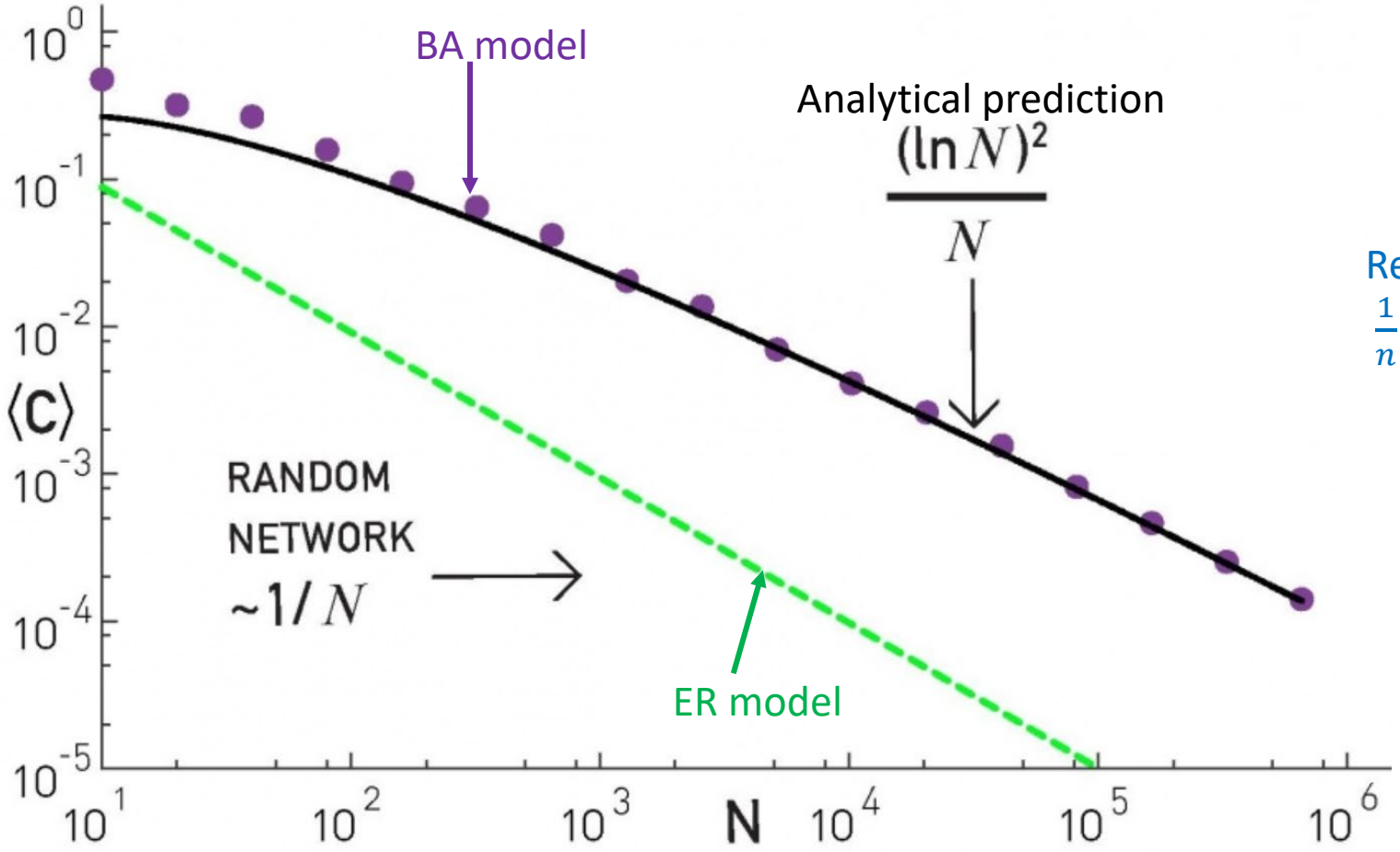


Clustering dependency

on degrees or size of network

ER & BA model: C_{ws} as a function of the network size (N)



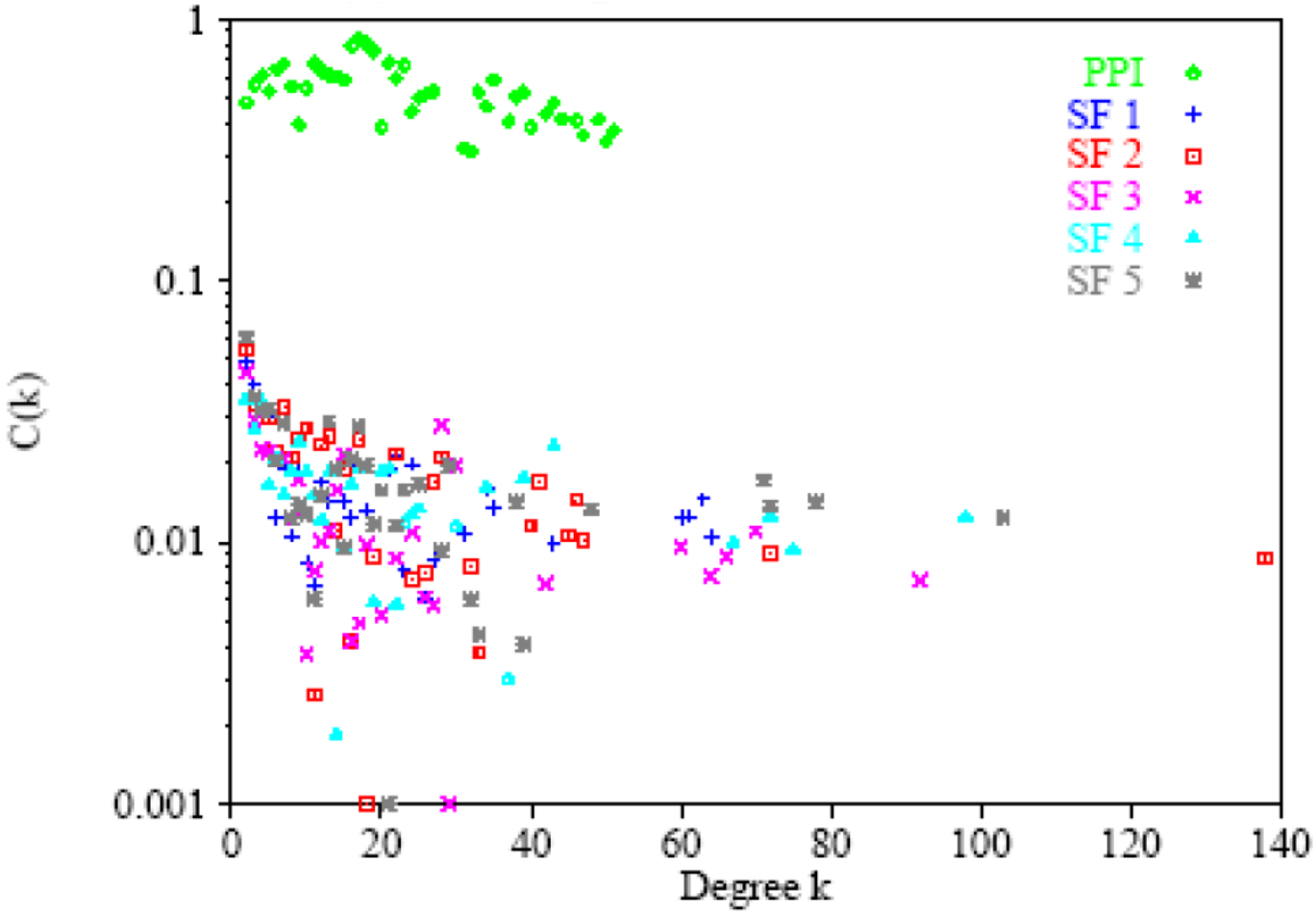


Recall: $C_{ws}(G) = \frac{1}{n}\sum_{i \in V(G)} C_i$

Barabási–Albert network is locally more clustered than a random network.

L. Barab´asi. http://barabasi.com/networksciencebook/chapter/5#diameter

$PPI\ and\ SF$: clustering as a function of degree k

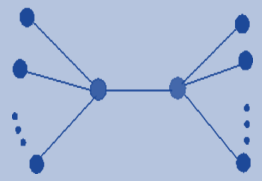


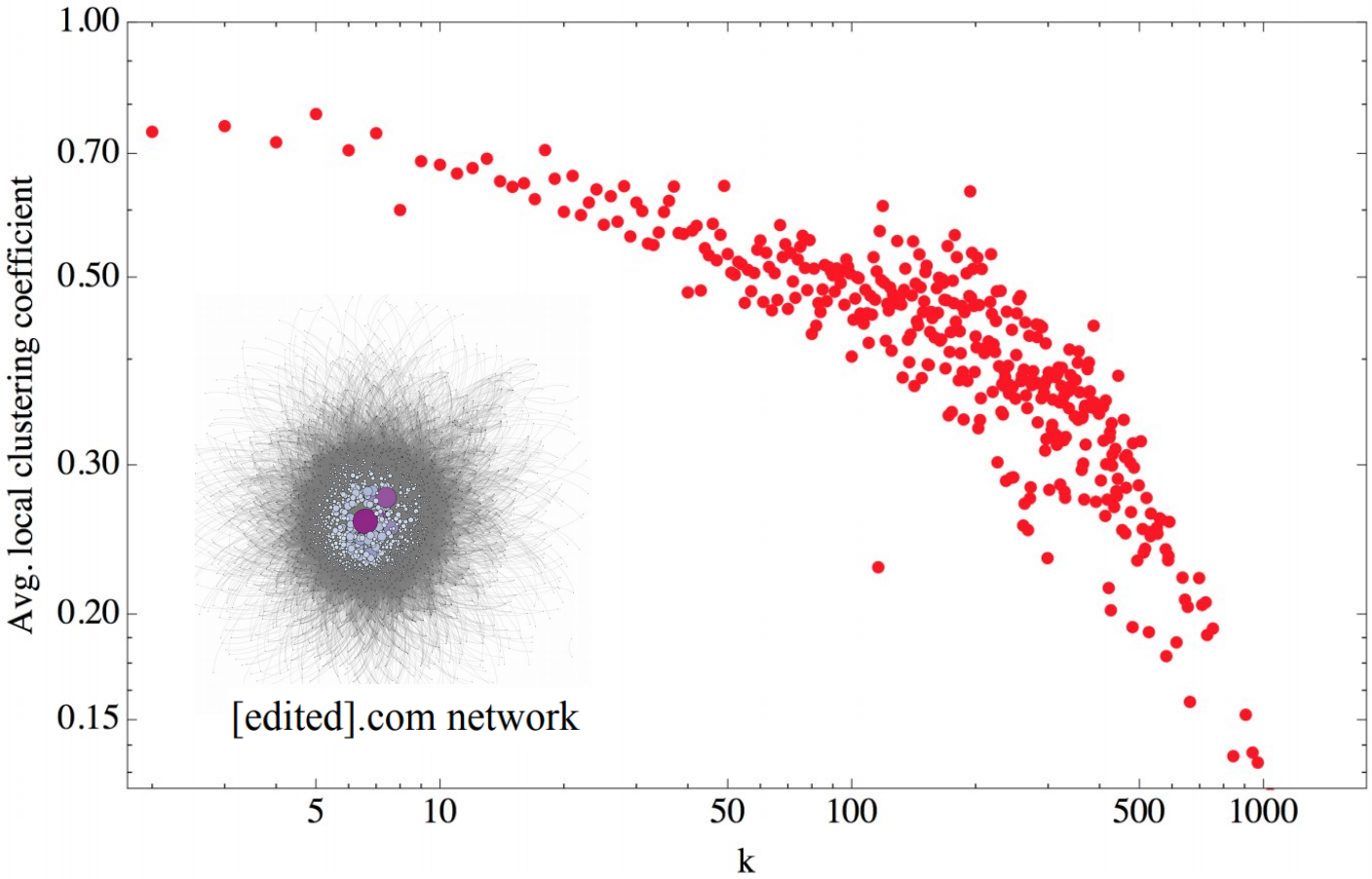
PPI: protein-protein interaction netw.

SF = scale free synthetic network.

Source: N. Prˆzulj, D. G. Corneil, and I. Jurisica. Modeling interactome: Scale free or geometric? arXiv:qbio. MN/0404017, 2004.

Online Social Networks: Avg Local Clustering as a function of degree





A node chosen at random has higher clustering than the one of a hub

Internet: Local Clustering Coefficient as a Function of Degree k

Notice: higher degree nodes exhibit lower local clustering coefficient

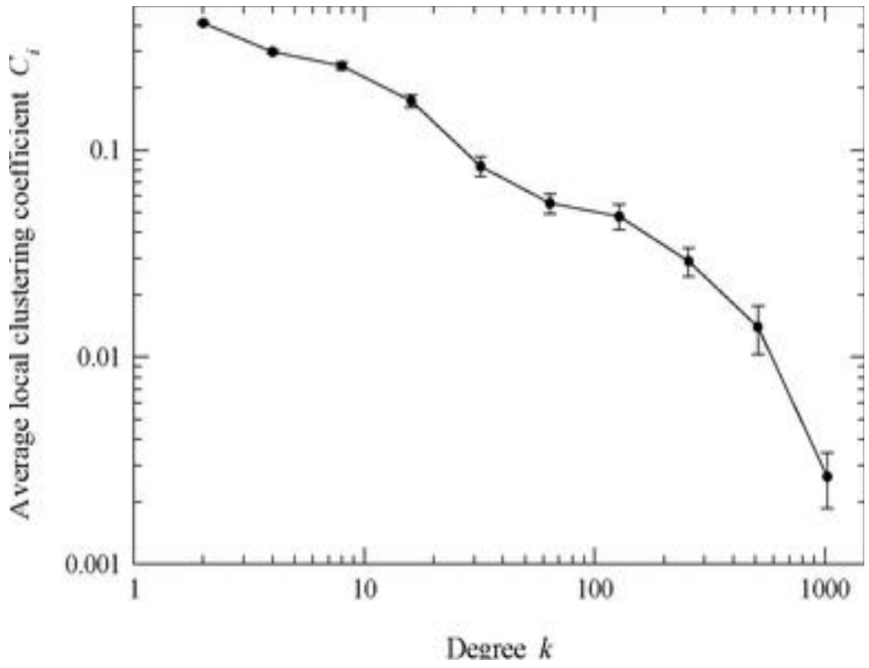
(with larger variance as well)

Thoughts on why?

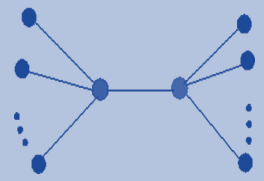
Example: Internet network at the Autonomous System level, averaged over all of the vertices of degree k. For nodes of degree k , the best fit is:

$$C_i\,(k) = k^{-\alpha},$$

where $.75 \leq \alpha \leq 1$



Local clustering coefficient



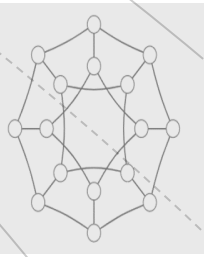
Possible explanations for the decrease in C_i as degree increases:

Vertices tend to group in communities, sharing mostly neighbors within the same community

Some vertices have small/large degree based on the size of the community

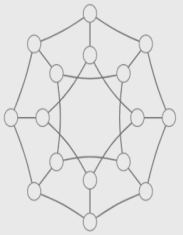
Smaller communities are denser → larger C_i

Communities may be connected by large degree nodes and being a connector will decrease its value of C_i of these large degree nodes.



Extension of Clustering Coefficient

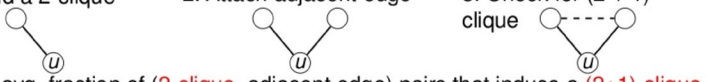
Extensions



Frequency of cliques in clique+edge occurrences:

Our higher-order view through clique expansion.

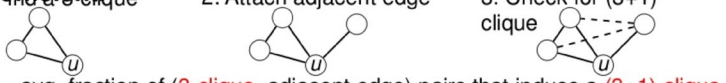
1. Find a 2-clique
2. Attach adjacent edge
3. Check for (2 + 1)-clique



$\bar{C}_2$ = avg. fraction of (2-clique, adjacent edge) pairs that induce a (2+1)-clique

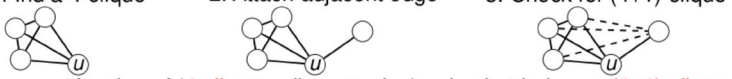
Increase clique size by 1 to get a higher-order clustering coefficient.

1. Find a 3-clique
2. Attach adjacent edge
3. Check for (3+1)-clique



$\bar{C}_3$ = avg. fraction of (3-clique, adjacent edge) pairs that induce a (3+1)-clique

1. Find a 4-clique
2. Attach adjacent edge
3. Check for (4+1)-clique



$\bar{C}_4$ = avg. fraction of (4-clique, adjacent edge) pairs that induce a (4+1)-clique

Local, average, and global higher-order clustering coefficients.

Third-order ***local*** clustering coefficient at node u.

$$C_3(u) = \frac{\#}{\#}$$

Third-order ***average*** clustering coefficient.

$$\bar{C}_3 = \frac{1}{n}\sum_u \frac{\#}{\#} = \frac{1}{n}\sum_u C_3(u)$$

Third-order ***global*** clustering coefficient.

$$C_3 = \frac{\sum_u \#}{\sum_u \#}$$

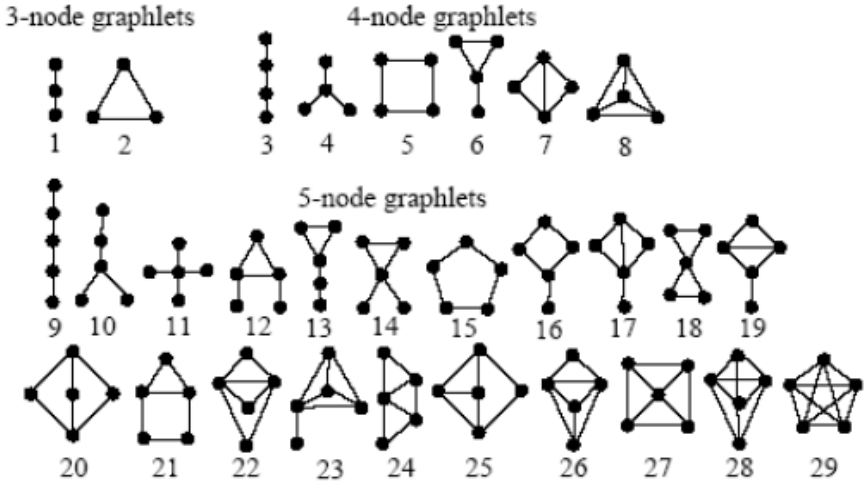
Benson, Yin, Leskovec, "Higher-order clustering coefficient" (2017)
https://www.slideshare.net/arbenson/higherorder-clustering-coefficients-80864022

Additional Extensions

Clustering coefficient measures the density of K_3 in networks

- Count the density of other motifs

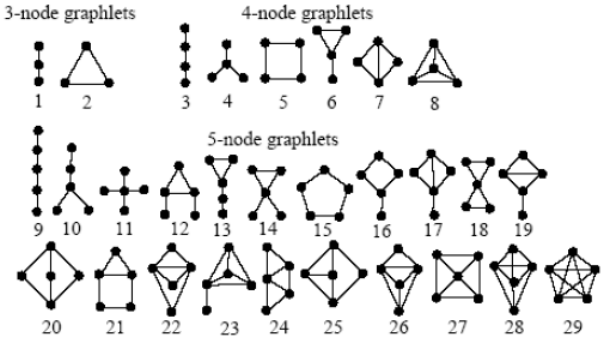
Enumeration of all graphs of fixed n:

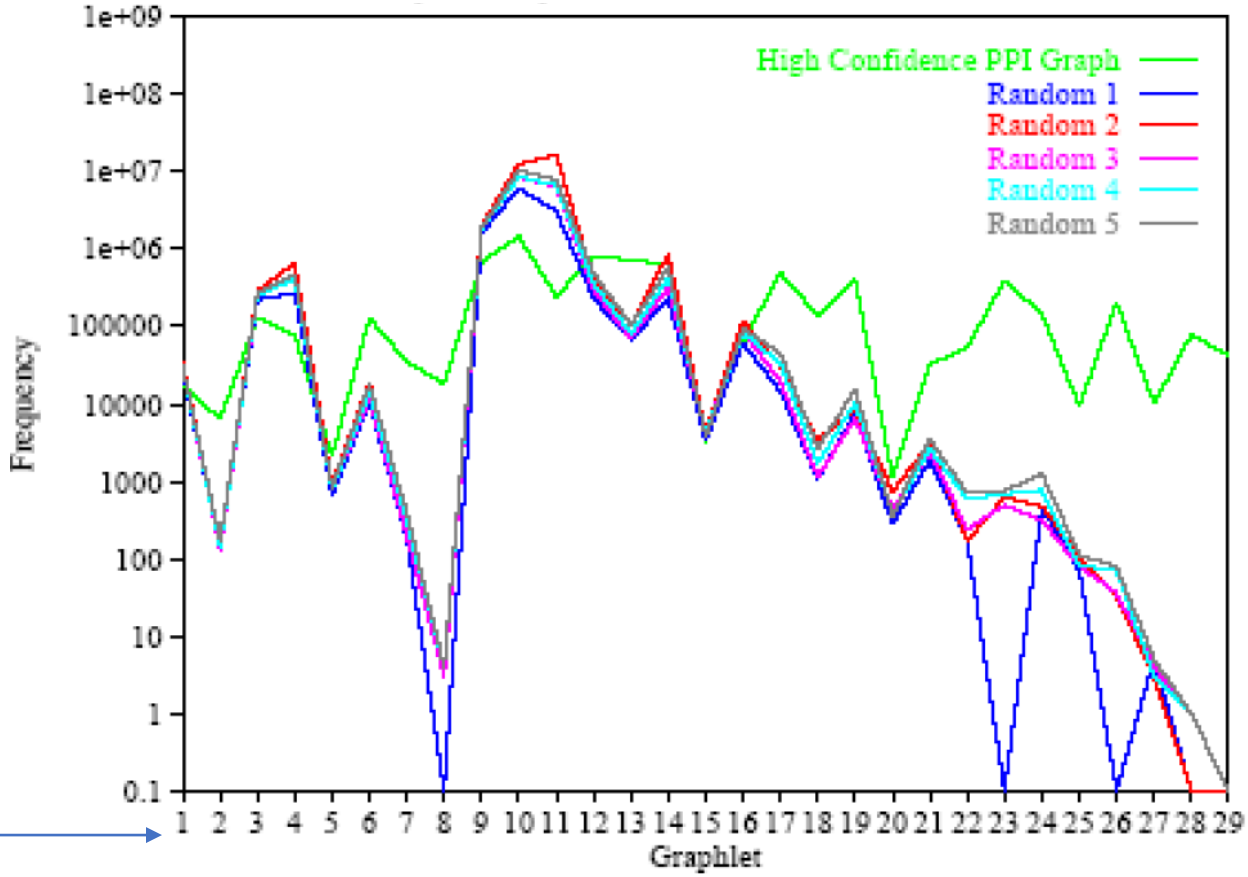


Source: N. Prˆzulj, D. G. Corneil, and I. Jurisica. Modeling interactome: Scale free or geometric? arXiv:qbio. MN/0404017, 2004.

Graphlet frequency in Scale Free network

Ordering of Graphlets





Source: N. Prˆzulj, D. G. Corneil, and I. Jurisica. Modeling interactome: Scale free or geometric? arXiv:qbio. MN/0404017, 2004.

Higher order clustering in WS networks

Local, average, and global higher-order clustering coefficients.

Third-order ***local*** clustering coefficient at node *u*.

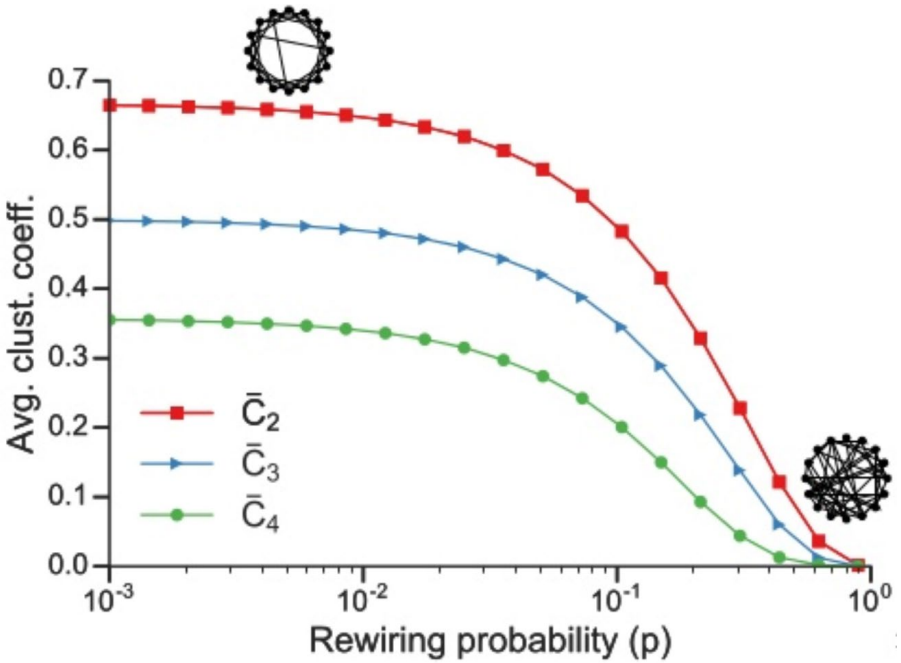
$$C_3(u) = \frac{\#\ \text{(triangles through } u\text{)}}{\#\ \text{(open wedges centered at } u\text{)}}$$

Third-order ***average*** clustering coefficient.

$$\bar{C}_3 = \frac{1}{n}\sum_u \frac{\#\ \text{(triangles through } u\text{)}}{\#\ \text{(open wedges centered at } u\text{)}} = \frac{1}{n}\sum_u C_3(u)$$

Third-order ***global*** clustering coefficient.

$$C_3 = \frac{\sum_u \#\ \text{(triangles through } u\text{)}}{\sum_u \#\ \text{(open wedges centered at } u\text{)}}$$



Benson, Yin, Leskovec, "Higher-order clustering coefficient" (2017)
https://www.slideshare.net/arbenson/higherorder-clustering-coefficients-80864022

References

- Newman, “The Structure and Function of Complex Networks” http://epubs.siam.org/doi/pdf/10.1137/S003614450342480
- Source: L. Barabasi. http://barabasi.com/networksciencebook/chapter/5#diameter
- Benson, Yin, Leskovec, “Higher-order clustering coefficient” (2017) https://www.slideshare.net/arbenson/higherorder-clustering-coefficients-80864022
- N. Prˆzulj, D. G. Corneil, and I. Jurisica. Modeling interactome: Scale free or geometric? arXiv:qbio. MN/0404017, 2004.