Computing Cliques and Cavities in Networks

Dinghua Shi[1]*, Zhifeng Chen[2], Xiang Sun[2], Qinghua Chen[2]*, Chuang Ma[3], Yang Lou[4], Guanrong Chen[4]*

([1]Department of Mathematics, College of Science, Shanghai University, China, shidh2012@sina.com;
[2]School of Mathematics and Statistics, Fujian Normal University, China, 920978196@qq.com;
[2]School of Mathematics and Statistics, Fujian Normal University, China, sun_xiang_3@163.com;
[2]School of Mathematics and Statistics, Fujian Normal University, China, qinghuachen@fjnu.edu.cn;
[3]Department of Internet Finance, School of Internet, Anhui University, China, chuang_m@126.com;
[4]Department of Electrical Engineering, City University of Hong Kong, China, felix.lou@my.cityu.edu.hk,
[4]Department of Electrical Engineering, City University of Hong Kong, China, eegchen@cityu.edu.hk)

Supplementary Information

1. k-Cores and Computable Networks

For the real USair, Jazz and Yeast networks[19], the number of cliques of different orders is limited to not more than 10^7 as detailed in Table SI-1.

Table SI-1 Number of cliques of different orders in real networks

Network	0-cliques	0-cliques	2-cliques	3-cliques	4-cliques
USAir	332	2126	12181	61072	243506
Jazz	198	2742	17899	78442	273697
Yeast	2375	11693	60689	424444	**2454474**

Network	5-cliques	6-cliques	7-cliques	8-cliques	9-cliques
USAir	766659	1931547	3947163	6608097	**9121594**
Jazz	845960	**2416059**			
Yeast					

If the number of cliques does not decrease with the increase of the order, it will become impossible to compute them by using personal computers.

It is noted that, in any network of a fixed size, except trees, its number of cliques of different orders has a peak value as the order number increases, namely it is first increasing and then decreasing. For instance, for a fully-connected network of size N, the numbers of its m-th order cliques are: $m_0 = C_m^1$, $m_1 = C_m^2$, $\cdots$, $m_{N-2} = C_N^{N-1}$, $m_{N-1} = C_N^N$, where it peaks at $\left(\frac{N}{2} - 1\right)$-clique (if N is even) or $\left(\frac{N-1}{2} - 1\right)$-clique (if N is odd). For example, when $N = 30$, it peaks at the 14-clique, with m_{14} =155117520; when $N = 25$, it peaks at the 12-clique, with $m_{12} = C_{25}^{13} = 5{,}200{,}300$.

Given limited computational resources, how can one determine if a given network is computable? For relatively large-scale and dense networks, k-core decomposition[17] may be used to roughly give an estimate. The k-core technique can be used to determine the cell of different orders, where all nodes on the k-sell have degree larger than or equal to k. The cell with the largest core value is the core of the network, where the connection is dense, therefore it can be used for measuring the order of the largest clique in the network. For example, in the Jazz network, the 29th cell has 30 nodes and 435 edges, implying that this is a fully-connected network; therefore, its core is a 29-clique, which is the order of the largest clique of the Jazz network. In the USAir network, the largest core value is 26, where the core has 35 nodes and 539 edges; therefore, its largest clique is a 21-clique, which is close to the core value 26. For the Yeast network, its core has 64 nodes and 1623 edges, which is known to have largest core value of 40; although the computation here reaches up to 6-clique, it can be seen that the order of the largest clique would not be small. The detailed core values of USAir, Jazz and Yeast are summarized in Table SI-2, where $\boldsymbol{m}_{\mathrm{i}}$ is the core value of the i-core, $i = 0, 1, 2, \dots, 29$.

Table SI-2 Core values of real networks

Core value	m_0	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9
USAir-26	35	539	4938	30580	137428	468604	1248988	2656044	4570650	6425067
Jazz-29	30	435	4060	27405	142506	593775	2035800	5852925	14307175	30045015
Yeast-40	64	1623	22344	196991	1222179	5656082	20278476			
Core value	m_{10}	m_{11}	m_{12}	m_{13}	m_{14}	m_{15}	m_{16}	m_{17}	m_{18}	m_{19}
USAir-26	7419660	7055424	5520504	3540415	1847164	774823	256755	65498	12370	1624
Jazz-29	54627300	86493225	119759850	145422675	155117520	145422675	119759850	86493225	54627300	30045015
Yeast-40										
Core value	m_{20}	m_{21}	m_{22}	m_{23}	m_{24}	m_{25}	m_{26}	m_{27}	m_{28}	m_{29}
USAir-26	132	5	0							
Jazz-29	14307175	5852925	2035800	593775	142506	27405	4060	435	30	1
Yeast-40										

The above analysis shows that, given the limited computational resources today, if the number of k-cliques is up to the order of 10^7 then the largest core value of the network should not be larger than 30, or even should be restricted to be below 25.

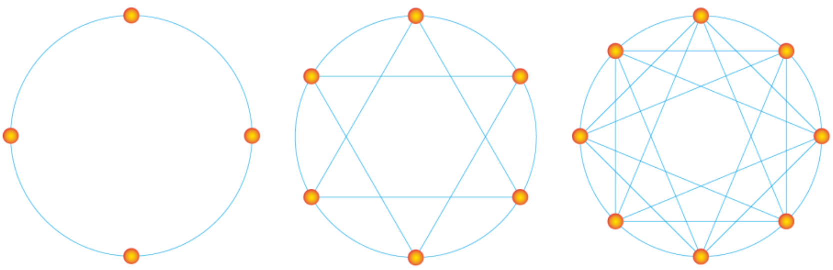
If the k-core decomposition[17] is performed by removing all nodes of degree $k = 1$ then some new nodes of degree $k \leq 1$ may emerge, and these nodes need to be removed as well, until all nodes have degree $k > 1$. All removed nodes and edges constitute 1-core with core value 1. This process continues for $= 2, 3, \ldots$, until the highest value k_{max} at which all nodes will be removed, and this last core has a core value k_{max}, and is the core of the original network.

The same idea can be used for cliques, named k-clique decomposition. Consider the sample network shown in Fig. 1, for instance. This network does not have 0-core and 1-core, and its 2-core contains nodes 6, 7, 8 and edges (3,6), (3,8), (6,7), (6,14), (7,8). Its 1-clique is composed of edges (3,6), (3,8), (6,7), (6,14), (7,8). Its 3rd sell consists of nodes 1, 2, 3, 4, 5 and edges (1,2), (1,3), (1,4), (1,5), (2,3), (2,4), (2,5), (3,4), (5,9). Its 2-clique is composed of edges (1,2,5), (9,10,11), (9,10,13), (9,11,12), (9,12,13), (10,11,14), (10,13,14), (11,12,14), (12,13,14). Its 4th cell consists of nodes 9, 10, 11, 12, 13, 14and edges (9,10), (9,11), (9,12), (9,13), (10,11), (10,13), (10,14), (11,12), (11,14), (12,13), (12,14), (13,14). Its 3-cliques is composed of (1, 2, 3, 4). This example shows the difference between the k-coe decomposition and the k-clique decomposition.

2. Smallest Possible Cavities of Different Orders

The concept of cavity comes from homology group in algebraic topology. Cavity is a special topological structure. The 1-cavity and 2-cavity have been found by observation[8]. In general, a smallest n-cavity is the smallest cycle consisting of some n-cliques, where the number of such n-cliques is larger than the number of boundaries of $(n+1)$-cliques. Furthermore, a smallest n-cavity can be obtained by introducing 2 more nodes, each connects to all nodes in the smallest $(n-1)$-cavity. Today, it is suspected that there is as high as 11th-order cavity in the neural network of the brain[9]. It is also known that the smallest k-cavity has a characteristic number[5]: $\chi = 1 + (-1)^k$.

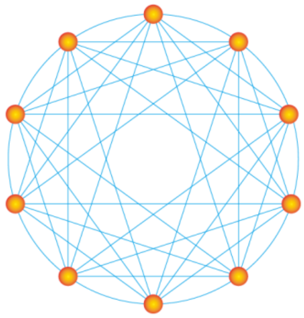
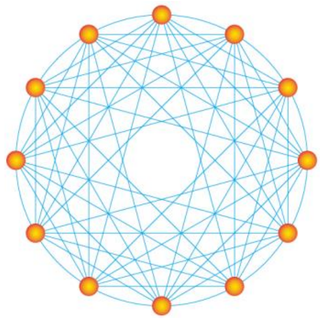
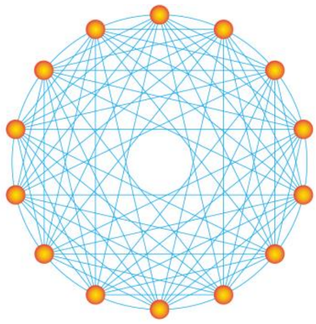
Numbers and features of smallest cavities of order 1 to order 11 are summarized in Fig. SI-1.



1-cavity: m_0=4， m_1=4； χ=0

2-cavity: m_0=6， m_1=12， m_2=8； χ=2

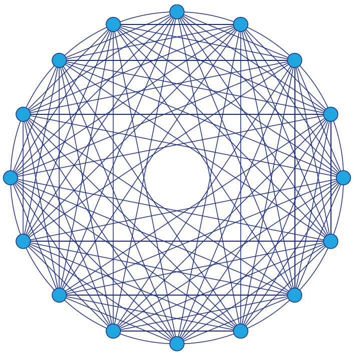
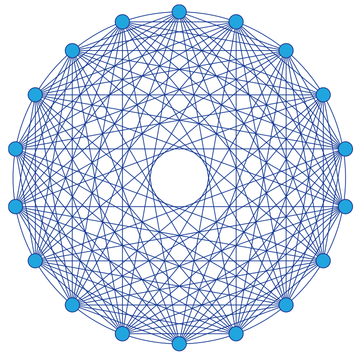
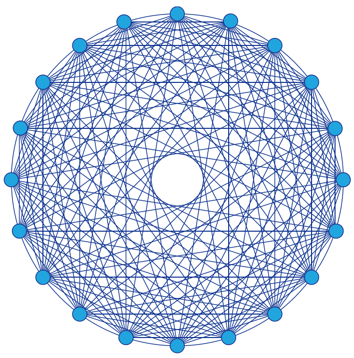
3-cavity: m_0=8， m_1=24， m_2=32， m_3=16； χ=0

4-cavity: m0=10，m1=40，m2=40，m3=80，m4=32；χ=2

5-vacity: m0=12，m1=60，m2=120，m3=240，m4=192，m5=64；χ=0

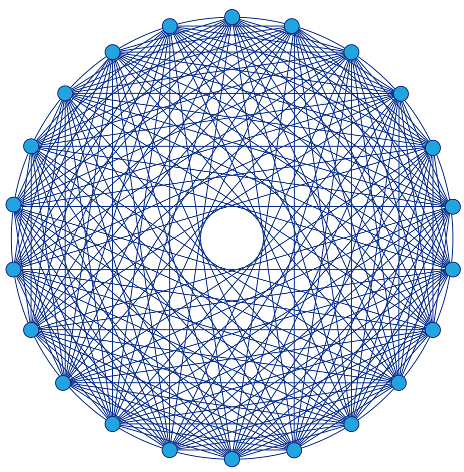
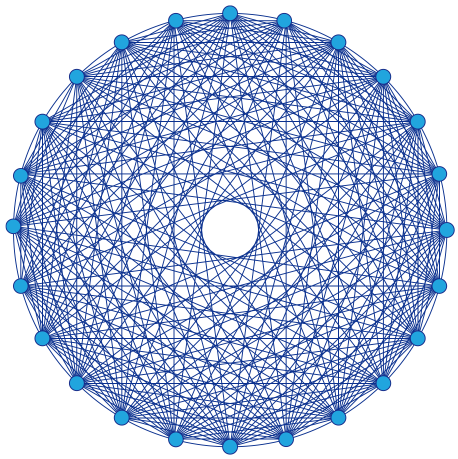
6-cavity: m0=14，m1=84，m2=280，m3=560，m4=672，m5=448，m6=128；χ=2

7-cavity: m_0=16，m_1=112，m_2=448，m_3=1120，m_4=1792，m_5=1792，m_6=1024，m_7=256；χ=0

8-cavity: m_0=18，m_1=144，m_2=672，m_3=2016，m_4=4032，m_5=5376，m_6=4608，m_7=2304，m_8=512；χ=2

9-cavity: m_0=20，m_1=180，m_2=960，m_3=560，m_4=3360，m_5=8064，m_6=13440，m_7=15360，m_8=11520，m_9=1024；χ=0

10-cavity: m_0=22，m_1=220，m_2=1320，m_3=5280，m_4=14784，m_5=29568，m_6=42240，m_7=42240，m_8=28160，m_9=11264，m_{10}=2048；χ=2

11-cavity: m_0=24, m_1=264, m_2=1760, m_3=7920, m_4=25344, m_5=59136, m_6=101376, m_7=125720, m_8=112640, m_9=67584, m_{10}=24576, m_{11}=4096; χ=0

Figure SI-1. Smallest cavities of order 1 to order 11.

3. Cliques and Cavities in C. elegans Network

For a dataset of C. elegans with 297 neurons and 2148 synapses[21], its cliques, ranks and the number of cavities are all obtained by using an available algorithm[9] and the proposed algorithm, with results summarized in Table SI-3.

Table SI-3. C. elegans Network

Clique	m_0	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8
	297	2146	3241	2010	801	240	40	2	0
Rank	r_0	r_1	r_2	r_3	r_4	r_5	r_6	r_7	r_8
	0	296	1713	1407	599	202	38	2	0
Betti number	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8
	1	139	121	4	0	0	0	0	0

Based on the data in Table SI-3, using the 0-1 programming, it is possible to find 4 different 3-cavities for details below.

The first 3-cavity with 8 nodes (85,13,3,164,163,119,118,158) is surrounded by the following 16 3-cliques:
(85,13,3,164) (13,3,164,163) (3,164,163,119) <u>(164,163,119,118)</u>
(163,119,118,158) (119,118,158,85) (118,158,85,13) (158,85,13,3)
(85,3,164,119) (119,158,85,3) (3,163,119,158) (158,13,3,163)
(163,118,158,13) (13,164,163,118) (118,85,13,164) (164,119,118,85)

The second 3-cavity with 11 nodes (163, 3, 162, 119, 154, 167, 118, 227, 85, 13, 164) is surrounded by the following 28 3-cliques:
(163,3,162,119) (3,162,119,154) (162,119,154,118) (119,154,118,167)
(154,118,167,13) (118,167,13,227) (167,13,227,3) (13,227,3,85)
(227,3,85,119) (3,85,119,164) (85,119,164,118) (119,164,118,163)
(118,163,119,162) (162,118,163,13) (13,162,118,154) (154,13,162,3)
(3,154,13,167) (167,3,154,119) (119,167,3,227) <u>(227,119,167,118)</u>
(118,227,119,85) (85,118,227,13) (13,85,118,164) (164,13,85,3)
(3,164,13,163) (163,3,164,119) (13,163,118,164) (163,3,162,13)

The third 3-cavity with 8 nodes (171,13,3,195,185,119,118,173) is surrounded by the following 16 3-cliques:

(171,13,3,195) (13,3,195,185) (3,195,185,119) (195,185,119,118)
(185,119,118,173) (119,118,173,171) (118,173,171,13) (173,171,13,3)
(171,3,195,119) (119,173,171,3) (3,185,119,173) (173,13,3,185)
(185,118,173,13) (13,195,185,118) (118,171,13,195) (195,119,118,171)

The fourth 3-cavity with 8 nodes (173,13,3,227,195,119,118,185) is surrounded by the following 16 3-cliques:

(173,13,3,227) (13,3,227,195) (3,227,195,119) (227,195,119,118)
(195,119,118,185) (119,118,185,173) (118,185,173,13) (185,173,13,3)
(173,3,227,119) (119,185,173,3) (3,195,119,185) (185,13,3,195)
(195,118,185,13) (13,227,195,118) (118,173,13,227) (227,119,118,173)

For 2-cavities, only those with eight nodes are listed here. A total of 4 were found, which are divided into two types, as shown in Figure SI-2.

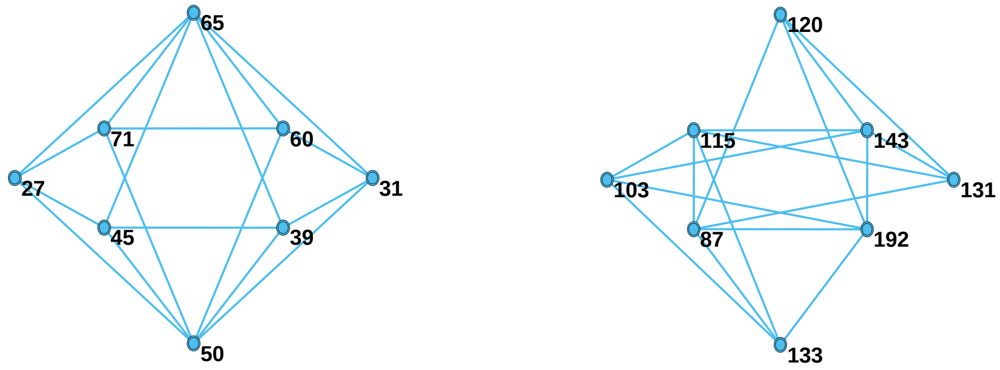


Figure SI-2. Two types of 2-cavities.

The first type of 2-cavity with 8 nodes (65, 31, 39, 45, 27, 71, 60, 50) is surrounded by the following 12 2-cliques:

(65,31,39) (65,39,45) (65,45,27) (65,27,71) (65,71,60) (65,60,31)
(50,31,39) (50,39,45) (50,45,27) (50,27,71) (50,71,60) (50,60,31)

The second type of 2-cavity with 8 nodes (120, 131, 143, 192, 87, 115, 103, 133) is surrounded by the following 12 2-cliques:

(120,131,143) (120,131,87) (120,192,87) (120,143,192)
(103,143,192) (103,115,143) (103,115,133) (103,133,192)
(87,133,192) (87,115,133) (87,115,131) (115,131,143)