

二部圖一個猜想

陳宏賓

一、二部圖

這裡我們要談的圖 (graph) 是指一個有序對 $G = (V, E)$, V 是一個有限集 (非空), 它所包含的元素稱為點 (vertex); E 是 V 中的點所形成的 (部份) 二元子集的集合, 意即 $E \subseteq \{e : e \subseteq V, |e| = 2\}$, E 中每一個二元子集 e 則稱為邊 (edge)。為了方便以見, 我們將 $e = \{u, v\}$ 記作 $e = uv$, 而 u 和 v 則稱為 e 的端點。在這裡, uv 和 vu 是沒有分別的。

圖的研究起始於十八世紀知名數學家歐拉 (Euler), 他將當時流傳在卡尼斯堡 (Königsburg) 的七橋問題清楚地用點邊描述並解決 [1], 這也是後來一般常見的益智遊戲「一筆劃」[2]; 關於圖研究的領域稱為圖論, 說到圖論, 最知名的莫過於「四色定理」(至今仍未有令人滿意的數學證明)。而第一本圖論的書是在相隔兩百年後的 1936 年才由卡內基 (König) 所撰寫 [3]。至今不到一百年間, 圖論的廣泛應用和研究動能相輔相成, 使其在近年來發展非常迅速。

一個圖稱為二部圖 (bipartite graph) 就是其點集 V 可以分成 A 和 B 兩部份, 而且每一條邊都恰有一端點在 A 另一端點在 B 。二部圖在圖論的研究中是一個很基本的圖類, 舉凡兩種事物之間關聯的表達都可以用二部圖來呈現, 因此受到相當大的關注。其中, 二部圖的匹配問題就是二十世紀初非常熱門的話題, 當年卡內基 (König) 的圖論一書有一大部份是在談論二部圖, 由此, 可見一斑。我們以下只考慮簡單圖 (simple graph), 即邊集 E 中所有元素 e 都不同且每一條邊 e 的兩個端點也都不同。

二、點數與邊數的關係

點數與邊數的關聯在圖論的研究上是個重要的議題。透過簡易的計算, 我們可以知道:

A. 任一 n 個點的圖最多有 $\frac{n(n-1)}{2}$ 條邊。

此達到最多邊數的圖則稱為完全圖 (complete graph), 即任意兩點都有一條邊相連。

也不難得到:

B. 任一 n 個點的二部圖最多有 $\lfloor \frac{n^2}{4} \rfloor$ 條邊。

達此最多邊數二部圖的 A, B 兩部份點數分別為 $\lfloor \frac{n}{2} \rfloor$ 和 $\lceil \frac{n}{2} \rceil$, 且 A 中所有點都和 B 中所有點相連, 記作 $K_{\lfloor \frac{n}{2} \rfloor, \lceil \frac{n}{2} \rceil}$ 。
這也是最多邊數的完全二部圖。

事實上, n 個點且「不含三角形」的最多邊數圖也就是這個圖 $K_{\lfloor \frac{n}{2} \rfloor, \lceil \frac{n}{2} \rceil}$ 。緬投 (Mantel) 在 1907 年證明了:

C. 任一 n 個點且不含三角形的圖 G 最多只有 $\lfloor \frac{n^2}{4} \rfloor$ 條邊, 且這個圖就是 $K_{\lfloor \frac{n}{2} \rfloor, \lceil \frac{n}{2} \rceil}$ 。

證明: 假設 $A \subseteq V(G)$ 是 G 中最大的獨立集 (independent set), 意即 A 的任意兩點都沒有邊。因為不含三角形, 所以任一點 x 的鄰居所成的集合必為獨立集, 因此, 我們知道 $d(x) \leq |A|$ (這裡 $d(x)$ 表示 x 的鄰居數量)。令 $B = V(G) \setminus A$ 。因為每一條邊至少有一個端點落在 B , 所以邊數

$$|E(G)| \leq \sum_{x \in B} d(x) \leq |A||B| \leq \left(\frac{|A| + |B|}{2}\right)^2 = \frac{n^2}{4},$$

由於 $|A|$ 和 $|B|$ 都是整數的關係, 我們可以更進一步推論 $|E(G)| \leq \lfloor \frac{n^2}{4} \rfloor$ 且等號成立時的圖就是 $K_{\lfloor \frac{n}{2} \rfloor, \lceil \frac{n}{2} \rceil}$ 。□

匈牙利的圖靈 (Turán) 在 1941 年給了上述定理一般性的推廣, 將之從不含 3 個點的完全圖 (也就是上述的三角形) 推廣至不含 r 個點的完全圖。

D. 任一 n 個點且不包含 r 點完全圖的圖最多只有 $\frac{(r-2)n^2}{2r-2}$ 條邊。

這種圖就是 n 個點的均勻完全多部圖, 意即有 n 個點均勻分成 $r-1$ 部, 任兩點在不同部有一條邊, 若在同一部內則沒有邊。後人為了感念圖靈的貢獻, 也稱作圖靈圖 (Turán graph)。

這類不含某種特定子圖的圖的最多邊數的研究是屬於極端圖 (extremal graph) 研究的範疇。極端圖研究的源頭始於匈牙利數學家圖靈, 並且在匈牙利開枝散葉, 就連著名的組合數學家艾狄胥 (Erdős) 也為其著迷, 著有相當多的研究成果。至今才 70 年左右的時間, 已經開展出一個專門的研究領域 – 極端圖理論 (Extremal Graph Theory), 有興趣的讀者可以參考 [4]。

三、張黃猜想

今天要談的, 不是極端的概念, 而是中庸!

以下稱 H 是 G 的導出子圖 (induced subgraph) 指的是, $V(H) \subseteq V(G)$ 且 $E(H) = \{uv \in E(G) : u, v \in V(H)\}$ 。

猜想 1 (張黃猜想 1980).¹ 任一 2^k 條邊的簡單二部圖, 必有一個恰好 2^{k-1} 條邊的導出子圖。

¹「張」是張鎮華教授, 目前任教於台灣大學數學系; 而「黃」是黃光明教授, 已經從交通大學應用數學系退休。

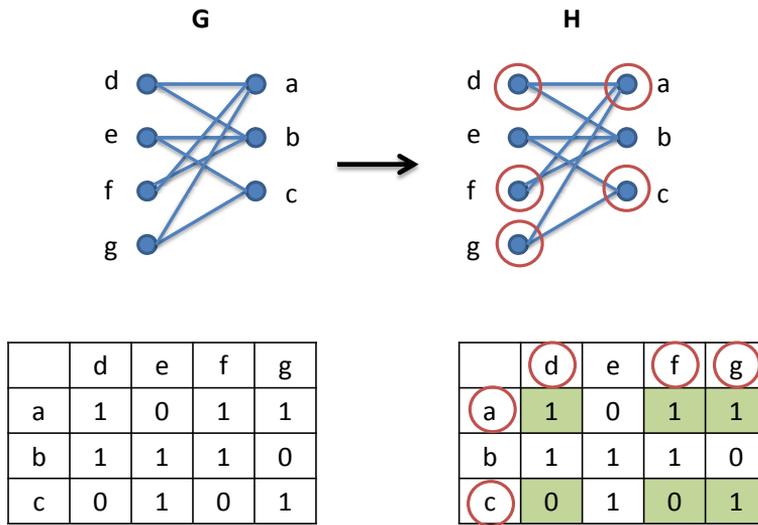


圖 1: 圖 G 為 7 個點 8 條邊的二部圖, 不難發現圖 H 為圖 G 的一個 4 條邊導出子圖, 其中 $V(H) = \{a, c, d, f, g\} \subseteq V(G)$ 且 $E(H) = \{ad, af, ag, cg\}$ 。

若將二部圖用連接矩陣來表示, 則張黃猜想有一個在矩陣上等價對應的版本。

猜想 2. 任一有 2^k 個 1 的 0-1 矩陣, 存在恰好 2^{k-1} 個 1 的子矩陣。

圖 1 是關於張黃猜想及其矩陣對應的一個例子。

如果將「簡單圖」限制拿掉, 那麼我們很容易可以找出反例, 如圖 2。

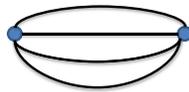


圖 2: 不存在 2 條邊的導出子圖。

更進一步, 若我們將「二部圖」限制拿走, 則同樣能夠找出反例, 如圖 3。

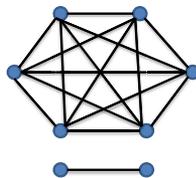


圖 3: 6 個點的完全圖加 1 條獨立的邊, 此圖共有 16 條邊, 但卻不存在 8 條邊的導出子圖。

最後, 「點數是 2 的正整數次方」這個條件也是重要的, 否則同樣存在反例, 如圖 4。

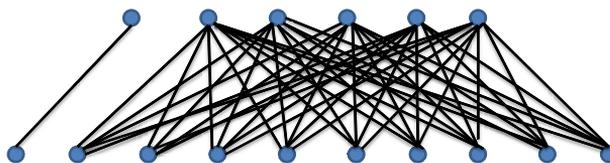


圖 4: 此為一個 46 條邊的簡單二部圖 $K_{5,9} \cup e$, 但卻不存在 23 條邊的導出子圖。

四、猜想緣起與發展

這個關於二部圖的猜想，來自一個看起來跟圖論毫無關係的新興領域 – 群試 (Group Testing)。

群試起源

時間倒回到 1941 年，地球上正經歷著砲火隆隆的第二次世界大戰，當時奉行孤立主義的美國還尚未參與戰爭。一直到了 12 月 7 日這天，一發意料之外的砲彈落在美國的珍珠港海軍基地，才打破了美國的寧靜。日軍這場偷襲不僅造成美軍傷亡慘重，更令世界局勢發生重大變化，美國羅斯福總統在發生日軍偷襲珍珠港事件後，旋即宣佈參戰。為參與歐洲及亞洲的戰場，美國要在短時間內招募大量士兵，而在當時梅毒這可怕又致命的傳染病猖獗，為免梅毒在士兵間傳染開來，進而影響戰力，徵召來的士兵都必須經過嚴格的血液篩檢，而梅毒的血液篩檢在當時又相當昂貴，美國軍方為此頭痛不已。

針對這個問題，在醫院工作的朶夫曼 (Dorfman) 提出了一套有效節省篩檢費用的方法 [5]，他建議與其將血液樣本一個個檢驗，何不先將若干份血液樣本混合，再檢驗此混合樣本是否遭受梅毒感染，如果呈現陰性反應(即未受感染)，則代表對應這些血液樣本的士兵們都未受感染；反之，如果呈現陽性反應(受感染)，則再將對應的這些血液樣本一個個檢驗，找出所有受梅毒感染的士兵。朶夫曼這個簡單的想法就是群試理論的原型。

基本模型

有 n 個血液樣本，用 0-1 向量 $\mathbf{x} = (x_1, \dots, x_n)$ 來表示 (0: 未受感染; 而 1: 被感染)，其中有 d 個受到感染，而目標是用少量的檢測 (group test) 將所有受感染的樣本找出。每次檢測的是任意子集合 $S \subseteq \{x_1, \dots, x_n\}$ ，得到的結果 $Q(S)$ 有 2 種可能: $Q(S) = 1$ 如果 S 中有被感染樣本; $Q(S) = 0$ 如果 S 中所有樣本都未被感染。

理論下界

令 $t(n, d)$ 表示此模型下所需的檢測次數。由於總共有 $\binom{n}{d} \geq \left(\frac{n}{d}\right)^d$ 種可能的分佈，且每次檢測只會得到 2 種可能，根據夏農 (Shannon) 所發展出來的資訊理論 (Information Theory)，我們知道 $2^{t(n,d)} \geq \binom{n}{d} \geq \left(\frac{n}{d}\right)^d$ ，兩邊同時取 \log 得到 $t(n, d) \geq d \log \frac{n}{d}$ 。這時候我們稱

$d \log \frac{n}{d}$ 為 $t(n, d)$ 的理論下界。

幸運的是，這個理論下界是可以達到的，指得是用 $cd \log \frac{n}{d}$ 個檢測 (c 是一個常數)。有非常多的論文探討關於這個常數 c 的值，不過這又是另外一個故事了，有興趣的讀者不妨自己動動腦，或者參考這本關於群試理論的專門書 [6]。

直覺不一定是對的

有天，張和黃在做研究的過程中，發現了一個出乎意料的結果，問題是這樣子的：

問題 1. 如果有兩個集合 A 和 B ， A 中有 5 個樣本， B 中有 3 個樣本，若已知 A 和 B 中都恰有 1 個受到感染，那麼需要幾次檢測才足夠找出這 2 個受感染樣本呢？

直覺上，因為 A 和 B 是不相交的兩個集合，將它們分開處理，分別找出受感染樣本似乎是最好的策略。用簡單的二元搜尋法，我們可以知道分別處理所需的檢測次數為

$$\lceil \log_2 5 \rceil + \lceil \log_2 3 \rceil = 3 + 2 = 5.$$

然而，事實並非如此。令 $A = \{a_1, a_2, a_3, a_4, a_5\}$ 且 $B = \{b_1, b_2, b_3\}$ 。張和黃的演算法是這樣子的：

- 先測 $S_1 = \{a_1, b_1\}$ 。
 - 如果 $Q(S_1) = 1$ ，則再測 $S_2 = \{a_1\}$ 。
 - * 如果 $Q(S_2) = 1$ ，則知 A 中受感染樣本為 a_1 ，另外 B 中只有 3 個樣本因此可以用 $\lceil \log_2 3 \rceil = 2$ 個檢測完成。
 - * 如果 $Q(S_2) = 0$ ，則知 B 中受感染樣本為 b_1 ，另外 A 中僅剩下 $\{a_2, a_3, a_4, a_5\}$ 共 4 個可疑樣本，因此，也能夠用 $\lceil \log_2 4 \rceil = 2$ 個檢測完成。
 - 如果 $Q(S_1) = 0$ ，則知 a_1 和 b_1 都未受感染， A 和 B 分別只剩下 4 個和 2 個可疑樣本，因此，能夠用 $\lceil \log_2 4 \rceil + \lceil \log_2 2 \rceil = 3$ 個檢測完成。

綜合以上，我們可以發現只要 4 次檢測就能夠找出這 2 個受感染樣本，比分開處理所需的 5 次檢測好一點，而事實上，4 次剛好是理論下界 $\lceil \log_2 5 \times 3 \rceil$ ，也是最好的可能了。

有了這個小例子的佐證，張和黃開始對這問題的一般性感到興趣，並且提出了：

問題 2. $\lceil \gamma \rceil$ 對任意不相交的兩集合 A 和 B ，是不是都存在只用理論下界 $\lceil \log_2 |A||B| \rceil$ 次數的檢測方法呢？

張黃猜想與問題 2 的關係

我們先試著將問題 2 中 $|A| \times |B|$ 種受感染樣本可能的分佈想成一個 $|A| \times |B|$ 的矩陣，每一個位置填上「*」表示未確定，填上「1」表示其對應行和列的兩個樣本即為受感染樣本，填「0」表示對應行和列的兩個樣本至少有一個是未受感染。

每一次的檢測只有兩種可能的結果，每一種結果會分別排除一些分佈的可能，意即，有些分佈會從 * 被斷定為 0，直到某一次檢測能夠判定哪一個分佈是 1 其他全部都是 0 為止。因此，如果存在一種檢測的策略，能夠保證在每一次檢測中，將矩陣中剩下未確定的位置數量減少一半，則表示經過 $\lceil \log_2 |A||B| \rceil$ 次檢測後，全部未確定的 * 都將變成 0 或 1，達成目的。

現在就用問題 1 這個例子，來解說它與張黃猜想的關係。想像一個 $V(G) = (A, B)$ 的完全二部圖 G ，它的連接矩陣有 $3 \times 5 = 15$ 條邊，在此先填入 15 個 *。首先我們找出一個 8 條邊的導出子圖 G_0 ，其點集為 $V(G_0) = \{b_2, b_3, a_2, a_3, a_4, a_5\}$ ，而定義另外 7 條邊所形成的圖為 G_1 ；其次，找一個包含於 G_0 的 4 條邊的導出子圖 G_{00} ，其點集為 $V(G_{00}) = \{b_3, a_2, a_3, a_4, a_5\}$ ，剩下的 4 條邊所成的圖定義為 G_{01} ；從 G_1 中找一個 4 條邊的導出子圖 G_{10} ，其點集為 $V(G_{10}) = \{b_1, a_2, a_3, a_4, a_5\}$ ，定義 G_1 中剩餘 3 條邊所形成的圖為 G_{11} ；依此一直下去直到邊數為 1，新增下標 0 代表其導出子圖，下標 1 為其剩餘圖，依序定義出這些圖：

$$\begin{aligned}
 G &= \begin{pmatrix} * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \end{pmatrix}, & G_0 &= \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & * & * & * & * \\ 0 & * & * & * & * \end{pmatrix}, & G_1 &= \begin{pmatrix} * & * & * & * & * \\ * & 0 & 0 & 0 & 0 \\ * & 0 & 0 & 0 & 0 \end{pmatrix}, & G_{00} &= \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & * & * & * & * \end{pmatrix}, \\
 G_{01} &= \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, & G_{10} &= \begin{pmatrix} 0 & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, & G_{11} &= \begin{pmatrix} * & 0 & 0 & 0 & 0 \\ * & 0 & 0 & 0 & 0 \\ * & 0 & 0 & 0 & 0 \end{pmatrix}, & G_{000} &= \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & * & * \end{pmatrix}, \\
 G_{001} &= \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & * & * & 0 & 0 \end{pmatrix}, & G_{010} &= \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & * & * \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, & G_{011} &= \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & * & * & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, & G_{100} &= \begin{pmatrix} 0 & 0 & 0 & * & * \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \\
 G_{101} &= \begin{pmatrix} 0 & * & * & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, & G_{110} &= \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ * & 0 & 0 & 0 & 0 \\ * & 0 & 0 & 0 & 0 \end{pmatrix}, & G_{111} &= \begin{pmatrix} * & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, & G_{0000} &= \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & * \end{pmatrix}, \\
 G_{0001} &= \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & * & 0 \end{pmatrix}, & G_{0010} &= \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & * & 0 & 0 \end{pmatrix}, & G_{0011} &= \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & * & 0 & 0 & 0 \end{pmatrix}, & G_{0100} &= \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & * \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix},
 \end{aligned}$$

若邊數 m 不是剛好為 2 的整數次方，如 $2^i \leq m \leq 2^{i+1}$ ，則視為增加一些虛擬的邊使其總數為 2^{i+1} ，以 2^i 為一半。

$$\begin{aligned}
 G_{0101} &= \begin{pmatrix} 00000 \\ 000*0 \\ 00000 \end{pmatrix}, G_{0110} = \begin{pmatrix} 00000 \\ 00*00 \\ 00000 \end{pmatrix}, G_{0111} = \begin{pmatrix} 00000 \\ 0*000 \\ 00000 \end{pmatrix}, G_{1000} = \begin{pmatrix} 0000* \\ 00000 \\ 00000 \end{pmatrix}, \\
 G_{1001} &= \begin{pmatrix} 000*0 \\ 00000 \\ 00000 \end{pmatrix}, G_{1010} = \begin{pmatrix} 00*00 \\ 00000 \\ 00000 \end{pmatrix}, G_{1011} = \begin{pmatrix} 0*000 \\ 00000 \\ 00000 \end{pmatrix}, G_{1100} = \begin{pmatrix} 00000 \\ 00000 \\ *0000 \end{pmatrix}, \\
 G_{1101} &= \begin{pmatrix} 00000 \\ *0000 \\ 00000 \end{pmatrix}.
 \end{aligned}$$

對照前面提到的演算法，第一次先測 $S_1 = \{a_1, b_1\}$ ，如果 $Q(S_1) = 0$ ，則表示 a_1 和 b_1 都是未受感染樣本，因此，剩下的可能用圖來表示的話，即為 G_0 ；反之如果 $Q(S_1) = 1$ ，則剩下的可能即為 G_1 。依此類推，我們不難發現這些圖都恰好分別對應到某一種檢測結果，一直到最後只剩下一條邊時，那條邊所代表的就是所求的兩個受感染樣本。

接下來的問題可能在於，找到一個導出子圖後，例如 G_0 ，我們該如何設計出一個檢測內容恰好符合這個導出子圖呢？其實很簡單，只要檢測所有不在 G_0 中的點即可，那就是 $\{a_1, b_1\}$ 。

經由上面的討論可以知道，證明張黃猜想即正面證明了問題 2。但不幸的是，反過來並不成立；也就是說，即使證明了問題 2 是對的，也不足以證明張黃猜想。事實上，問題 2 早在 1981 年(也就是提出後的隔年)就被張和黃證明出來 [8]。雖然原始的問題已經被解決，但張黃猜想卻開啓了圖群試理論的新潮流：

給定一個二部圖 $G = (V, E)$ ，已知其中有一條邊是壞邊其餘都是好邊，試問如何用最少的檢測次數將壞邊找出來？這裡的檢測仍然可以是任意的點集合 $S \subseteq V$ ，但 $Q(S)$ 則是不同於傳統群試： $Q(S) = 1$ 表示點集合 S 所生成的子圖中有一條邊是壞邊， $Q(S) = 0$ 則表示生成的子圖中所有邊都是好邊。

自此，圖上的群試理論的研究正式展開。有興趣的讀者，可以參考 [9, 10] 及裡面豐富的參考資料。針對張黃猜想的研究結果並不多，目前最好的結果記載於暨南大學阮夙姿和台灣大學張鎮華教授合著的這一篇論文 [11]，證明了

定理 1 ([11])。對於任意 2^k 條邊的二部圖， $k \leq 5$ ，張黃猜想都是對的。

參考資料

1. L. Euler, Solutio problematics ad geometriam situs pertinentis, *Commentarii Academiae Scientiarum Impericalis Petorplictae*, Vol. 8 (1736), pp. 128-140.
2. 徐力行, 沒有數字的數學, 天下文化, 2003年。
3. D. König, *Theory of Finite and Infinite Graphs*, translated by R. Mcloart with commentary by W. T. Tutte, Biskhäuser, Boston, 1990.

4. B. Bollobás, *Extremal Graph Theory*, New York: Dover Publications, 2004.
5. R. Dorfman, The detection of defective members of large populations. *Ann. Math. Statist.*, Vol. 14, No. 4, pp. 436-440, 1943.
6. D. Z. Du and F. K. Hwang, *Combinatorial Group Testing and Its Applications*, New Jersey, World Scientific, 1993.
7. G.J. Chang and F.K. Hwang, A group testing problem, *SIAM J. Alg. Disc. Meth.*, Vol. 1, No. 1, pp. 21-24, 1980.
8. G.J. Chang and F.K. Hwang, A group testing problem on two disjoint sets, *SIAM J. Alg. Disc. Meth.*, Vol. 2, No. 1, pp. 35-38, 1981.
9. M. Aigner, *Combinatorial Search*, John Wiley and Sons, New York, 1988.
10. M. Bouvel, V. Grebinski and G. Kucherov, Combinatorial search on graphs motivated by bioinformatics applications: a brief survey. *Lect. Notes Comput. Sci.*, Vol. 3787, pp. 16-27, 2005.
11. J.S. Juan and G.J. Chang, Group testing in graphs, *J. Comb. Optim.*, Vol. 14, pp. 113-119, 2007.

—本文作者任職中央研究院數學研究所—

小品

小數點

作者：唐翊寧

我是仰望實數的小數點，站立在數大的人群裡。
 只是默默地凝視旁人，是安靜又卑微的配角。
 在浩瀚塵世中，我是多麼的渺小，
 但，當需要我在整數和小數的中間畫上分界點時，
 那時的我是不能缺少的主角。

—本文作者為中國科技大學財務金融系學生—