



FACEBOOK

究竟知道什麼

在這家公司深處，一組社會科學領域的專家正在尋找關乎人類行為、前所未聞的洞見。他們的發現，可能促使 Facebook 以新方法利用會員們的資料——並且顛覆我們對社會如何運作的理解。

撰文 湯姆·席蒙奈特 Tom Simonite ·
《科技評論》(Technology Review) 的資深資訊科技編輯

本刊獨家取得美國麻省理工學院

Technology Review 期刊圖文授權

Technology Review, Published by MIT.

TECHNOLOGY REVIEW

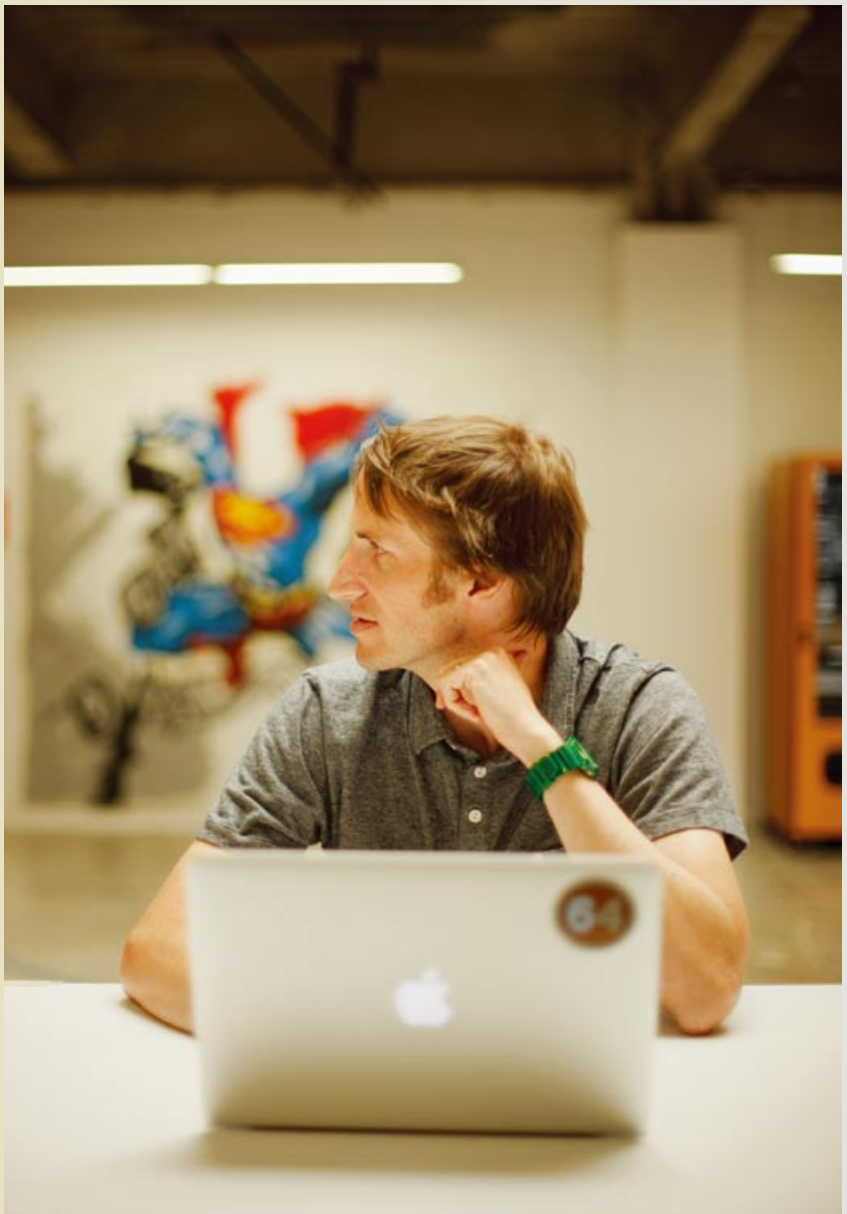
Internet URL: www.technologyreview.com

Technology
PUBLISHED BY MIT
Review

Facebook 創辦人馬克·祖克柏 (Mark Zuckerberg) 曾經語帶自負地打趣說，如果 Facebook 是個國家，九億會員會使它成為第三大國。臉書鉅細靡遺記錄會員生活的點點滴滴，所下的功夫遠遠超越過去或現在的任何政權。私人談話、家庭照片，以及道路旅行的紀錄、生日、婚姻和死亡，全都流進這家公司的伺服器，並且留在那裡。Facebook 收集有史以來最廣泛的人類社交行為的資料，你的一些個人資訊，很可能是其中的一部分。

不過，即使 Facebook 已經融入現代生活，卻還沒有真正運用它對你我的所知去執行夠多的革新。現在這家公司的股票已經公開發行，必須開發新利潤的壓力，可能迫使它利用儲存的資訊，去做更多事情。那些隱藏的資料，隱約就像一團超大的影子，讓今天的線上廣告業務相形見絀，隱私意識強的網路使用者和 Google 等競爭對手也感到憂慮。每個人都有一種感覺，覺得這個前所未見的資源，將化為一個很大的衝擊，只是沒人確切知道那將會是什麼。

凱莫隆·馬洛 (Cameron Marlow) 領導 Facebook 的一支團隊，努力探索能從我們的所有資料知道什麼。這位 35 歲的高個子，不久前還坐在離祖克柏只有幾呎的地方。馬洛主管的小組，



馬洛自稱是 Facebook 的「內部社會學家」，他和組內同事基本上可以分析臉書收集的全部資訊。

不像臉書創辦人以及它的特色業務那樣吸引大眾的注意。這個團隊的內部名稱叫做「資料科學小組 (Data Science Team)」，有點像是社交網路時代的貝爾實驗室 (Bell Labs)。這個小組有十二個研究員——但今年人數可望增為兩倍。他們利用數學、程式設計技巧和社會科學，從我們的資料

中挖掘出新洞見，希望藉以增進 Facebook 的業務和擴張社會科學的範疇。臉書的其他分析師將研究重心放在和特定線上活動有關的資訊，馬洛的團隊卻能優遊於幾乎全部的臉書會員個資之海。可能包括 Facebook 的領導人在內，當屬這些研究員擁有最好的機會，能在這麼多的個人資訊中，



巴克希透過實驗，了解Facebook的使用者分享連結的方式，好讓他的團隊研究這座網站是否像個回音箱。

洞察出未知的事物。

Facebook能擁有這樣龐大的資訊，是因為它找到聰明的方法，在人們上網交友的同時，收集他們的資料。使用者會填寫年齡、性別、電子郵件地址等基本資料；有些人還會給更多額外的詳細資訊，例如感情狀態和行動電話號碼。去年秋天，臉書重新設計以

時間軸的形式呈現個人資料頁，讓使用者加進他們住在哪裡和工作地點等歷史資訊。在網站上分享的訊息和照片，往往標記精確的地點，而且過去兩年，Facebook利用「讚」(Like)鈕這種叫人上癮的發明，開始追蹤網際網路其他地方的活動。「讚」出現在微型運用程式和Facebook之外的其

他網站，網友只要按一下滑鼠，就能表示他們對某個品牌、產品或一段數位內容感興趣。去年秋天以來，當使用者在某些微型應用程式(apps)或網站，聽一首歌或看一條新聞，即使沒有按「讚」，這個資訊也會傳到臉書，自動蒐集使用者的站外線上生活。這項功能推出的前五個月，Facebook收集到會員利用網路聽歌超過五十億次。把這些資訊和會員在站內的交友關係地圖結合起來，你會擁有人們極其豐富的生活和互動紀錄。

「這是世界上首次看到這種規模和品質的人類溝通資料，」馬洛帶著他特有的嚴肅表情說，等想到他能用資料做什麼事，又綻出一抹笑容。馬洛深信，探索這個資料庫，會使社會科學產生革命性的變化，讓我們更了解人為什麼表現出這些外顯行為。他的小組也能協助臉書影響人們的社交行為，往對它和它的廣告主有利的方向發展。這項研究甚至可能幫助臉書找到全新的獲利模式。

傳染性資訊

馬洛穿著一襲禮服襯衫外加牛仔褲，不同於和臉書其他許多員工，像學程式設計的大學生那樣，穿兜帽連身衣或T恤。在5月間臉書股票首次公開發行前不久，他和我在會議室見面。會議室的玻璃牆上，有祖克柏愛犬的六呎高漫

畫噴漆。他進來的時候，看起來比較像是年輕教授，不像學生。要是他在事業生涯之初，沒有發現網路公司有最豐富的人類互動資料，或許真的會選擇去教書。

2001年，馬洛在麻省理工學院的媒體實驗室(Media Lab)攻讀博士學程時，設立了Blogdex網站，會自動列出部落格散播的最具「傳染性」資訊。雖然這只是一項研究計畫，但很快就有很多人瀏覽，導致馬洛的伺服器當機。Blogdex推出時，正值部落格爆炸性成長，網友趨之若鶩，但在部落格數量激增後，網友終於覺得被資訊壓垮，因此而有後來的Digg和Reddit等整合網站跟著出現。但是馬洛建立Blogdex的目的，不只是協助網路使用者追蹤網路上什麼資訊最熱門。Blogdex作為一種科學工具，用以發掘正在網路上成型的社交網路，並且研究它們如何散布一些想法。馬洛接著到雅虎(Yahoo)的實驗室，研究線上社交兩年。2007年，他進入Facebook服務，認為這是世界上研究人類社會最強而有力的工具。「有史以來首次，」馬洛說，「我們有了一具顯微鏡，不只允許我們在非常細微的層次，探討以前見不到的社會行為，也讓我們執行有數百萬使用者參與的實驗。」

馬洛的團隊和Facebook各單位的經理人合作，尋找他們可能用

得上的型態。舉例來說，他們研究一項新功能如何在社交網路的使用者之間散播開來。他們協助你找到可能認識，但尚未「加為朋友」的會員，並且點出你可能想要指定為「點頭之交」的人，讓泛泛之交的近況更新不出現在那麼顯著的位置。可是在軟體工程師有如搖滾明星，並把「快步邁前，突破現狀」掛在嘴上的一家公司內，這個團隊的存在顯得有點奇特。和資料小組的人共進午餐，你會覺得好像是和高等學府的大學生聚在一塊；除了平常的熟面孔，有時會加進新科博士，或者剛進入學校擔任教職的人。他們談如何對社會科學有所貢獻，多於如何推廣Facebook的產品或這家公司。團隊成員有的學過社會科學或心理學，有人則念電腦科學，後來開始將它用於研究人類行為。他們可以自由運用一些時間和Facebook的資料，探討人類行為的基本型態和動機，並將論文發表在學術期刊——就像貝爾實驗室的研究人員，在AT&T同時推進技術和基礎物理學那樣。

這家公司才八歲，還不曾擁有證明可行的商業模式，卻敢於養一支帶有這麼濃厚學術味的團隊，乍看之下似乎很奇怪，但馬洛表示，這是有道理的。「Facebook必須解決的最大挑戰，和社會學面對的挑戰相同，」他說。挑戰包括了解為什麼某些觀念會從少數

人流行起來，流傳到最後隨處可見，而其他人卻做不到這件事；或者一個人未來的行動，在多大的程度內是過去與朋友互動的產物。他補充說，發表論文並與大學研究人員團隊合作，將有助於Facebook改善它的產品。

Facebook可作為探討整個社會的縮影，理由之一在於1967年發表的最有名的真實世界研究，幾百名參與人嚐試將明信片寄給一位在波士頓的股東，證明地球上任何一個人，和其他任何一個人之間，都只隔著六層人脈(six degrees of separation)。Facebook和米蘭大學的研究人員合力進行研究，範圍包含迄2011年5月止的整座社交網路，人數相當於全球人口的10%以上。分析7億2100萬人的690億個朋友關係，結果顯示世界比我們所想的要小：通常只經過四個中間朋友，就能把任何人介紹給隨機一位陌生人。該篇論文簡潔有力地作結：「就世界上任何一個人來說，通常你朋友的朋友，認識他們朋友中的一位朋友。」這個結果也許不能擴延到地球上的每個人，但有好理由相信它和資料科學小組的其他研究，確實反映了Facebook站外的生活。去年普優研究中心(Pew Research Center)的網際網路與美國人生活計畫(Internet & American Life Project)發現Facebook的朋友，有93%已經見

過面。馬洛的一位研究員根據 Facebook 的活動，記錄表達正面或負面情緒的字或詞，來計算一個國家的「國民幸福毛額」(gross national happiness)。從詞彙量數的波動，可以看出國民幸福毛額相當準確：假日期間，它會躍升；受推崇的公共人物死亡時則下挫。2010年2月智利發生大地震，這個國家的分數急速下降，過了好幾個月才恢復。這次事件，似乎使智利全國在2011年3月日本發生大地震，以及後來的海嘯肆虐期間，表現得更具同情心；智利的國民幸福毛額下挫，但其他國家(日本除外)的數字並沒有波動。製作這個指數的亞當·克萊默(Adam Kramer)說，他的目的是要指出 Facebook 的資料，能以便宜且準確的方式追蹤社會趨勢——這對經濟學家和其他的研究員可能相當有用。

這支團隊發表的其他研究，對 Facebook 的策略有較明顯的用處，包括鼓勵人們以這座網站為生活重心，然後利用它所知的資訊銷售廣告。一份早期的研究，觀察朋友的近況更新類型，會鼓勵這座網站的新會員增添本身的資訊。今年的情人節前不久，資料科學小組貼出的一篇部落格文章，列舉使用者最近在 Facebook 上表示他們進入某種感情狀態或離開某種感情狀態時最常聽的歌曲。這可以幫助 Facebook 對使用

者的行為做出的預測——知道這件事，可能有助於它猜對你此時此刻較容易接受的廣告。剛結束一段感情的人，或許對民歌專輯比較感興趣。抑或一個人在遭逢朋友不幸死亡、情緒爆發的時刻，網頁不該亮出和他們的情緒有關的品牌。今天最有價值的線上廣告，是和若干網路搜尋結果並列的廣告，因為搜尋者已經準確表達他們想要什麼。這是為什麼 Google 的營業收入，是 Facebook 十倍的原因。但 Facebook 最後甚至可能在人們意識到前，就猜出他們想要或不想要什麼。

最近，資料科學小組開始利用它獨特的地位，微調並實驗 Facebook 的運作方式——就像科學家戳螞蟻窩那樣——看看使用者會有什麼樣的反應。艾坦·巴克希(Eytan Bakshy)曾經是密西根大學的博士生，和馬洛合作過，去年加入 Facebook。他想測試臉書好友是否創造出一具「回音箱」，令人們已聽聞的新聞與輿論更為發酵。因此他混合 2.5 億使用者的 Facebook 運作方式，七個星期內，記錄使用者彼此分享的 7,600 萬個連結。接著 Facebook 在隨機選定的 2.19 億個組合，不讓某個人看到某位朋友分享的連結。利用隱藏連結產生一個控制組，好讓巴克希評估使用者在有類似資訊來源和興趣的情況下，最後多常貼相同的連結。

他發現親密朋友強烈影響我們分享的資訊，但整體而言，他們的影響力和許多關係較遠的人——社會學家稱之為「弱關係」——合起來的影響力相比，顯得小巫見大巫。

弱關係的多樣化集合，有最強的力量決定我們看到什麼資訊。

這項研究提出強烈的證據，推翻困擾許多人的一個觀念：社會網路化產生了有害的「過濾氣泡」(filter bubbles)。這是積極行動主義者伊萊·帕理澤(Eli Pariser)用的詞，藉以說明我們為了符合自身的期望，調整所接收的網路資訊造成的影響。但這項研究也揭露 Facebook 擁有的力量，「如果每個人都看到 Facebook 的新聞遞送(News Feed)，而且控制資訊的傳播方式，那麼它就控制了資訊向社會的揭露方式，這是我們需要非常注意的事情，」馬洛說。他指出，他的團隊協助 Facebook 了解它對社會做了什麼，並且發表它的研究發現，以履行透明化的公共義務。另一項最近的研究，調查哪些種類的 Facebook 活動，讓人覺得獲得朋友更大的支持，也屬於同一類。

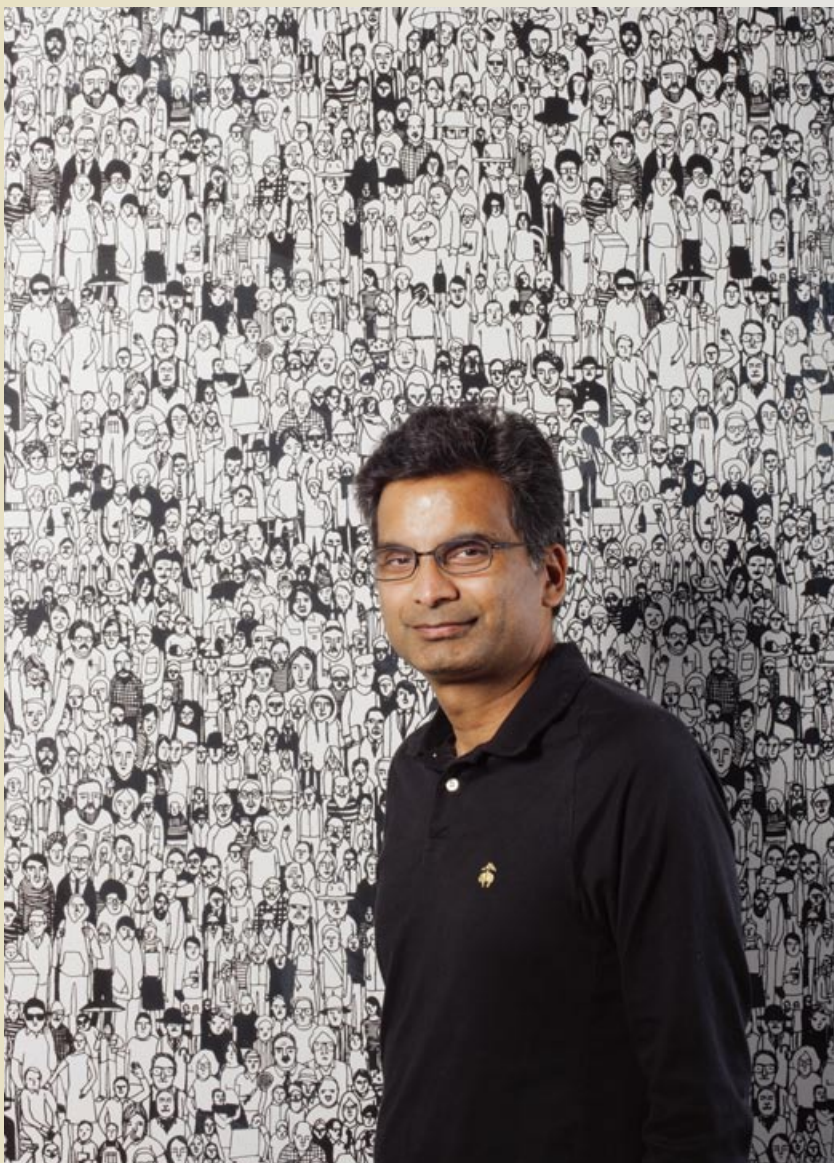
但馬洛畢竟任職於靠迎合廣告主的商業公司，他們想要控制使用者之間的資訊傳遞，讓公司業務欣欣向榮。巴克希正與資料科學小組之外的經理人合作，從社會影響力實驗的結果，找出和

廣告有關的發現。「廣告主和品牌也是臉書的一部分，因此讓他們在某種程度內知道人們如何分享他們產生的內容，是這個商業模式非常核心的部分，」馬洛說。Facebook在股票首次公開上市前告訴可能的投資人：看到朋友支持網站上的廣告，使用者記住廣告的可能性會高出50%。研究影響力如何運作，可以使廣告更為難忘，或者幫助Facebook找到一些方法，促使更多人分享或點按它的廣告。

浩大的社會工程

馬洛說，他的團隊想抽絲剝繭找出線上社交生活的規則，以了解Facebook社群裡發生了什麼事，而不是發展操控社群的方式。「我們的目標不是改變社會溝通的型態。」他說，「而是了解它，好讓我們調整我們的平台去適應，給使用者他們想要的體驗。」但從他的團隊所做的一些研究，以及Facebook領導人的態度，看得出這家公司並不排斥使用它的平台，微調使用者的行為。Facebook的員工和學術界的社會學家不同，有捷徑能將構想付諸實驗，而且受測人數高達數億。

四月間，因為和唸醫科的妻子共進晚餐的一席話，祖克柏決定利用Facebook的社會影響力，來提高器官捐贈的登記人數。於是Facebook讓使用者有機會在他們



艾加瓦爾正設想各種方法，以利Facebook管理龐大的資料寶庫——因此為臉書創造獨特且寶貴的技術能力。

的動態時報(Timeline)網頁上，勾選一個方格，表示他們已登記捐贈器官，進而發出通知給朋友。這個新功能引發如瀑布般層層而下的社會影響力，44個州的器官捐贈登記人數增加23倍。

馬洛的團隊正在發表美國上次期中選舉的結果，又給了一個引人注目的例子，讓我們看到Facebook引導使用者彼此影響

的潛力。自2008年起，臉書設計讓使用者表示他們已投票，並鼓勵他們通知朋友，說服朋友也應該去投票。2010年的選舉，馬洛的團隊比對了選民登記名冊和資料，觀察Facebook的哪些使用者受到推促之後，實際去投了票。

(馬洛強調，研究員用的是加密過的「匿名化」資料，無法比對特定的使用者和他們的投票紀

錄。)

這只是起步，Facebook曉得小小的變動，可以如何改變使用者的站外行為之後，終有一天「可能允許其他人以相同的方式使用Facebook，」馬洛說。舉例來說，如果美國心臟協會（American Heart Association）想要鼓勵吃得健康，或許可以參考Facebook社會工程的劇本。「我們希望成為一座平台，讓其他人用於發起變革，」他說。

廣告主也急於知道更深入的細節，什麼因素會使Facebook上的廣告宣傳影響人們在外部世界的行為，即使他們知道確實引導人的行為有其極限。「社會學是否會像工程學建造橋樑的方式那樣，我還不清楚，」鄧肯·華茨（Duncan Watts）說。華茨在微軟（Microsoft）最近設立的紐約研究實驗室研究運算社會學，曾和馬洛在雅虎的實驗室共事。「不過，如果你有足夠的資料，做出的預測會比完全胡亂猜測要好，而這才是真正有利可圖的地方。」

當資料倍增...

Facebook和Twitter和其他社群網路公司一樣，不曾享有Google等網際網路開路先鋒那樣的技術創新美名。如果矽谷是一所中學，那麼Google算是安靜的數學天才，不擅長交際應酬，卻發明出不可或缺的東西。Facebook就像惱人

的孩子，設立了一個具有社交動力的俱樂部，不管人們喜不喜歡，都必須加入。事實上，Facebook已經從Google和其他的數學天才公司挖角一大群才華洋溢的軟體工程師，打造並維持它那難以抗拒的俱樂部。Facebook的營運規模，使它發明出的硬體和軟體，令想要調整自我、適應「大資料」世界的其他公司艷羨不已。

就像傳遞技術接力棒那樣，Facebook以擴大開放源碼軟體Hadoop效能的方式，建立起自己的資料儲存系統。Hadoop是受到Google的啟發，並在雅虎打造的軟體。Hadoop能使看似不可能執行的運算任務——例如處理Facebook使用者託交給它的所有資料，將它們分散到一個資料中心內部的許多伺服器上，讓運算順利進行。在設計Hadoop時，Facebook要將它用在這個目的上，所以沒有將資料科學放在心上，而採用專業且笨拙的程式設計，因此工程師發明Hive。Hive是現在獨立於Facebook的開放源碼軟體，為其他許多公司使用。

Hive就像一個翻譯服務，能以相當簡單的程式碼，來搜尋龐大的Hadoop資料庫。為了縮減運算上的需求，它可以要求從整個資料庫取得隨機樣本，這對被資料淹沒的公司來說極為寶貴。Facebook的工程單位主管薩米特·艾加瓦爾（Sameet Agarwal）

說，不少資料存在一個Hadoop資料庫中，大小超過100拍位元組（petabytes；相當於100萬吉位元組〔gigabytes〕）。艾加瓦爾負責資料設備，需要處理的資料數量正以指數級數成長。「幾年來，我們每年的資料數量增加一倍以上。」他表示，他的團隊必須不斷建立效率更高的系統。

馬洛的前任主管傑夫·哈默巴赫爾（Jeff Hammerbacher）開始發展Facebook本身的資料儲存與分析技術。他說，這些給了Facebook獨特的技術能力水準。

（2008年他離開Facebook，創立Cloudera公司，開發以Hadoop為基礎的系統，管理大量的資料集。）

大部分大型企業花很多錢，請甲骨文（Oracle）等老牌軟體公司提供資料分析和儲存服務。但是現在，哈默巴赫爾說，大公司試著了解Facebook如何在開放源碼系統上，處理龐大的資訊寶庫。「我最近在富達（Fidelity）待了一天，協助他們了解Facebook的『資料科學專家』是怎麼設置的...和其他無數公司也討論過相同的事，」他說。

由於每個產業的高階主管都想利用「海量數據」中的機會，他們對Facebook的資料技術表示強烈的興趣，所以它的廣告業務，和遠比廣告更有價值的未知可能性相比，可能只算小兒科。臉書開發用

於處理大量資訊的工具和技術，本身可能成為一種產品。

開採黃金

Facebook 需要新的收入來源，才能滿足投資人的期望。即使股票首次公開發行的表現令人失望，它的本益比還是高得驚人，無法靠這座網站現在顯示的無數便宜廣告撐起來。Facebook 接替昇陽電腦 (Sun Microsystems) 進駐位在加州門洛帕克 (Menlo Park) 的園區，3,500 位員工在可容納 6,600 人使用的空間中忙碌著。我走過一棟到處是空辦公桌的大樓；隔壁棟大樓空無一人。這附近有一塊空地，恐怕是要等某位研究員找到使用我們資料的方法，臉書才覺得值得砸錢去開發，顯而易見其財務壓力。

一種可能的商務模式很簡單，就是銷售從資訊中挖掘出來的洞見。創業投資公司葛雷洛克合夥 (Greylock Partners) 的常駐資料科學專家帕帝爾 (DJ Patil)，以前是 LinkedIn 資料科學小組的領導人，相信 Facebook 可以從 Google 的 AdSense 廣告業務發明者吉爾·艾勒巴茲 (Gil Elbaz) 那裡得到靈感。AdSense 帶來的收入占 Google 的四分之一以上。吉爾已經離開廣告業務，現在經營一家成長快速的新創公司 Factual，向企業收費，讓它們取用審慎管理的大量資料集。

臉書從免費的公共來源，以及購買私人的資料集，而收集到餐廳地點、名流的身體質量指數等種種資料；Factual 清理資料之後，將結果送上網際網路，作為隨選即用的知識庫，供軟體而非群眾使用。顧客用這些資訊來填補本身資料的缺口，以及製作更為聰明的 App 或服務；例如 Facebook 就使用 Factual 的商業地點資訊。帕帝爾指出，Facebook 本身可以成為資料來源，銷售因編纂使用者行為而取得的資訊給他人使用。他說，這些資訊可能是任何業務 (例如網路紅娘或流行音樂榜) 的基礎。假使 Facebook 能在

會顯示 Facebook 上如何分享各個網站的內容。當 Facebook 把它的「讚」鈕追蹤整個網路活動、人口統計或人們在網站上閱讀什麼的資訊都加進來，對企業的用處將提高許多。付費提供這種資料分析已有先例，2011 年底，Google 開始提供加值版本的服務，分析一家企業的網路流量，每年收費 15 萬美元。

回頭談 Facebook，馬洛不是決定公司收費提供什麼服務的人，但他的工作會影響那些決策。他說，不管發生什麼事，他的團隊的首要目標，都是照顧提供資料給 Facebook 的使用者之利益，並

Facebook 將銷售從客戶資訊中挖掘出來的洞見，
協助客戶用這些資訊來填補本身的缺口，
用來製作更聰明的 App 或服務。

不令使用者和管理機關反感的情形下走到這一步，將會非常賺錢。例如線上商店如果想要針對目標客戶展開促銷活動，可以付費使用 Facebook 的知識，得知哪些品牌在哪些地點最暢銷，或者若干產品一年當中的銷路起伏情形。

哈默巴赫爾爾相信 Facebook 有可能銷售它的資料科學，並且直指它目前免費供應給廣告主和網站擁有者使用的 Insights 服務，也許可以成為金雞母。Insights 服務

且運用那些資料，促使服務更加聰明。他說，這個過程中，他和同事會促進人類對自身的了解。這番話附和了祖克柏經常遭人懷疑，但似乎是真心的信念，他深信 Facebook 的使命是改善全世界社群的溝通。但請不要問到底要怎麼做，「很難預測我們會走到哪裡，因為我們處於這門科學的非常早期階段，」馬洛說，「畢竟 Facebook 的資料可能做到的事，多不勝數。」