Math and the police force

An algorithm discovered by an 18th century mathematician may be used to find patterns in the communication records of criminals, says s ananthanarayanan

There is no branch of detective science which is so important and so much neglected as the art of tracing footsteps.
— Sherlock Holmes in *A Study in Scarlet*

University of Alberta researcher Ray Patterson found that a mathematical algorithm known as the Steiner tree could reveal social-network connections used to perpetrate fraud.

THE classic picture at the scene of a crime is of the police measuring distances, dusting for fingerprints and questioning the neighbours. This last activity, questioning neighbours, is to map the social network, the points of contact, which are relative to the crime. In Enid Blyton books and also in Connan Doyle's, the social picture unfolds the pathway to the crime, though, in Doyle offerings, the path is one that readers did not notice.

Researcher Ray Patterson at the University of Alberta, Canada, and colleagues have reported in the journal, *Computer Fraud and Security*, that a mathematical algorithm discovered by 18th century mathematician Jakob Steiner could reduce the time taken to investigate present day cases of fraud. Related things are interrelated and the relationships form networks. Examples are the places in a city – these are connected by road, either directly or via other places, and often through more than one path. Some of the connections may be one-way streets, which means the path cannot be "retraced" and each path would have a capacity of what traffic it can carry.

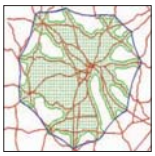
Another kind of network is a family tree, showing connections between persons through a common ancestor or through other relationships. A vast network of our times, of course, is the Internet, a network of computers that provides abundant alternate paths from any one computer to any other.

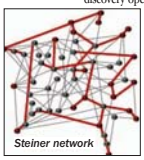
Nature is also built of networks, evolutionary trees, predator relationships, food cycles. And so are human relations, the family network, the friends' circle, professional relationships (either one's own on the other person's — which leads to the HIV trail) and, of late, socialising through the Internet.

While stable and efficient networks, as of roads, sea and air routes or supply chains, grow naturally and shape the growth of cities or industry, aggressive planning, development and competitiveness of recent times has called for the design of man-made networks for efficiency and economy. Mathematical studies of the basic issues have found ample application and, with the computing power that is available today, optimising the way of working is routine if one has to stay in business.

A classic network problem is the *traveling salesman* one – the salesman has to visit a number of clients or cities and return, with the least distance, or it could be time, or cost of travel. The same solutions would work for the layout of components of microchips or in studies of molecular biology and there has been academic interest in efficiently finding solutions. The subject appears to have been formally addressed first by the mathematician-physicist WR Hamilton, who used the idea that nature followed the most economical path, to solve problems in mechanics. The routing and scheduling problem, in fact, fast becomes hellishly difficult, with each node affecting all other nodes, resulting in a mathematical entity called the *polynomial*, with as many variables, like n_1, n_2, n_3, etc, as nodes. The *brute force* method, of trying out all solutions, may be impossible and clever strategies to reduce labour have been developed. Between 1950 and 1980, there has been great progress and there are now exact solutions of problems that cover thousands of cities.

One kind of network problem is the *Steiner tree problem*, which is to interconnect a set of points with the shortest total length of connections. The problem permits introducing additional points, called Steiner vertices, like hubs that connect the spokes of a wheel, to economise the total length. The same idea is used by airlines for routing and pricing services. Other fields where the principles are used are in the design of integrated circuits, computer networks, layout of depots for distribution of products, even decision-taking procedures in organisational hierarchies. Jakob Steiner, the first to work on this pattern of relationships, was a German mathematician of the early 19th century and he made important contributions in many areas of mathematics.





Steiner network

Crime and fraud

Mathematical methods are used for planning how to position police personnel and resources for faster deployment. Another application is for identifying areas of rising crime and the regions where it is most productive to intensify patrolling. Ray Patterson of Alberta and colleagues have found that analysis of the network of contacts of fraudsters could be an area where the Steiner tree could be effective. Getting back to the image of police questioning neighbours to ascertain the points of contact with a crime, a similar approach in cases of present day frauds would be to trace the electronic connections that the players in the case maintain. But in the present day, with computer communication and mobile phones, the neighbourhood has expanded and there may be hundreds of contacts.

As it is, the police needs to take permission and clearance to open computer or telephone records. The work, thereafter, to analyse the information and identify leads to follow takes great time before anything useful is identified. Patterson and his colleagues explored how networks such as phone calls, business partnerships and family relationships are actually used to reach the essential relationships in a fraud investigation. When these relationships are layered, a pattern of connection can be recognised. Once unnecessary links are removed and false leads are extracted, the remaining connections are most likely the best suspects.

Patterson says that finding the shortest connection between the criminals and the crime then reduces to the form of the *Steiner tree*. "All of these things that we see in life, behind them is a mathematical representation," he says. "There are many, many different algorithms that we can pull off a shelf and apply to real-life problems."

The work he and his group have been doing is to recognise that the relationships that arise in a fraud investigation form a Steiner tree. This discovery opens up the scope of using the Steiner tree algorithm for analysis of the relationship data that the police collects in these cases. A course for further investigation could then be computer generated directly after digital data of computer social networks, emails and mobile calls are fed in.

There is already software that is used by law enforcement agencies, but this largely to classify and retrieve crime records, analysis of crime trends, etc.

The use of mathematics and software based on algorithms to filter out less promising avenues and optimise the line of investigation would save substantial time. "If you can reduce your legwork by even 20 per cent, that has massive manpower implications. I think algorithms like this one could help you reduce your legwork a lot more than that," Patterson says.

The writer can be contacted at simplescience@gmail.com

Reservoirs of liquid light

Computer models may identify many more habitable planets, writes sam masters

THERE are many more habitable planets in distant solar systems than previously thought. Research from scientists at the University of Aberdeen was presented at the British Science Festival recently showing the new models that help identify planets with a greater likelihood of having underground water and, therefore, able to support life.

Previous estimates of planets capable of supporting life had been based on the likelihood of there being surface water available. It was thought that for water to exist in liquid form, a planet must be a certain distance from its sun, in a "habitable" or "Goldilocks" zone.

Sean McMahon.

"Life as we know it requires liquid water. Traditionally, planets have been considered habitable if they are in the Goldilocks zone," said Sean McMahon of the University of Aberdeen's School of Geosciences.

"They need to be not too close to their sun but also not too far away for liquid water to persist, rather than boiling or freezing, on the surface. However, we now know that many micro-organisms reside deep in the rocky crust of the planet, not on the surface. There will be several times more (habitable) planets," he said.

The new research is based on estimates that even when the surface of a planet is frozen, huge quantities of water, teeming with life, could exist below, warmed by heat generated within the planet and not its star.

"There is a significant habitat for micro-organisms below the surface of the earth, extending down several kilometres," said Professor John Parnell of Aberdeen University. "And some believe that the bulk of life on earth could reside in this deep biosphere."

John Parnell.

The university is developing models to predict which planets might harbour underground reservoirs of liquid water with the possibility of alien life.

Suns warm planet surfaces but heat also comes from planet interiors. Crust temperature increases with depth so planets that are too cold for liquid water on the surface may be sufficiently warm underground to support life.

McMahon said, "We have developed a new model to show how 'Goldilocks zones'can be calculated for underground water and hence life. Our model shows that habitable planets could be much more widespread than previously thought."

The Independent, London

A question of mutation

Oncogenes, writes tapan kumar maitra**, are those whose presence can trigger the development of cancer and proto-oncogenes are converted into oncogenes by several distinct mechanisms**

SOME oncogenes are introduced into cells by cancer-causing viruses, whereas others arise by mutation from normal cellular genes. In either case, oncogenes code for proteins that stimulate excessive cell proliferation and/or promote cell survival. The first oncogene to be discovered was in the *Rous sarcoma* virus. Because this virus has only four genes, it was relatively easy to determine which gene triggers the development of cancer.

Mutational studies revealed that defects in one particular gene, called the *src oncogene*, yield mutant viruses that no longer cause cancer yet can still infect cells and reproduce normally. In other words, a functional copy of the src gene must be present before the virus can cause cancer. Similar approaches have led to the identification of several dozen other viral oncogenes. In cancers not caused by viruses, the first evidence for the existence of oncogenes came from studies in which DNA isolated from human bladder cancer cells was introduced into a strain of cultured mouse cells called 3T3 cells. The DNA was administered under conditions that stimulate transfection — that is, the uptake of the foreign DNA into the cells and its incorporation into their chromosomes. When transfected with the cancer cell DNA, some of the mouse 3T3 cells proliferated excessively; when these cells were injected back into mice, the animals developed cancer. Scientists therefore suspected that a human gene taken up by the mouse cells had caused the cancer.

To confirm the suspicion, gene cloning techniques were applied to DNA isolated from the mouse cancer cells. This resulted in the identification of the first human oncogene: a mutant RAS gene coding for an abnormal form of Ras, the protein. (Note that the names of human genes — for example, "RAS" are generally written in italicized capital letters, while the names of the proteins they produce, for example "Ras", are written without italics and often with only an initial capital letter.)

RAS was just the first of dozens of human oncogenes to be discovered. While these are defined as genes that can cause cancer, a single oncogene is usually not sufficient. For example, in the transfection experiments described in the preceding paragraph, introducing the RAS oncogene caused cancer only because the mouse 3T3 cells used in these studies already possessed a mutation in another cell-cycle control gene. If freshly isolated normal mouse cells are used instead of 3T3 cells, introducing the RAS oncogene by itself will not cause cancer. This observation illustrates an important principle: multiple mutations are usually required to convert a normal cell into a cancer one.

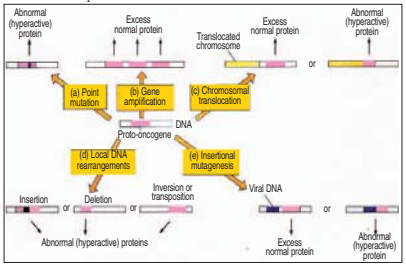
How do human cancers, most of which are not caused by viruses, come to acquire oncogenes? The answer is that oncogenes arise by mutation from normal cellular genes called *proto-oncogenes*. In spite of their harmful sounding name, proto-oncogenes are not bad genes that are simply waiting for an opportunity to foster the development of cancer. Rather, they are normal cellular genes that make essential contributions to the regulation of cell growth and survival. The term "proto-oncogene" simply implies that if and when the structure or activity of a proto-oncogene is disrupted by certain kinds of mutations, the mutant form of the gene can cause cancer. The mutations that convert proto-oncogenes into oncogenes are created through several distinct mechanisms, which are briefly described below. Chart here

Some of the oncogenes produced by these mechanisms code for abnormal proteins, whereas others produce normal proteins in excessive amounts, namely:

- Point mutation involves a single nucleotide substitution that creates an oncogene coding for an abnormal protein differing in a single amino acid from the normal protein produced by the proto-oncogene;
- Gene amplification creates multiple gene copies that are actively expressed, thereby producing excessive amounts of a normal protein;
- Chromosomal translocations move chromosome segments from one chromosome to another. This may either fuse two genes together to form an oncogene coding for an abnormal protein or it may place a proto-oncogene next to a highly active gene, thereby inducing the translocated proto-oncogene to become more active;
- Local DNA rearrangements (such as insertions, deletions, inversions and transpositions) can disrupt the structure of proto-oncogenes and cause them to produce abnormal proteins; and
- Insertional mutagenesis is triggered by the integration of viral DNA into a host chromosome near a proto-oncogene. This may either disrupt the proto-oncogene, thereby creating an oncogene coding for an abnormal protein, or it may enhance expression of the proto-oncogene and cause it to produce too much protein.

Point mutation: The simplest mechanism for converting a proto-oncogene into an oncogene is a point mutation — that is, a single nucleotide substitution in DNA that causes a single amino acid substitution in the protein encoded by the normal proto-oncogene. The most frequently encountered oncogenes of this type are the RAS oncogenes that code for abnormal forms of the Ras protein. Point mutations create abnormal, hyperactive forms of the Ras protein that cause the Ras pathway to be continually activated, thereby leading to excessive cell proliferation. RAS oncogenes have been detected in several human cancers, including those of the bladder, lung, colon, pancreas, and thyroid. A point mutation can occur at any of several different sites within a RAS oncogene, and the particular site involved appears to be influenced by the carcinogen that caused it

Gene amplification: The second mechanism for creating oncogenes utilises gene amplification to increase the number of copies of a proto-oncogene. When the number of gene copies is increased, it causes the protein encoded by the proto-oncogene to be produced in excessive amounts, although the protein itself is normal. For example, about 25 per cent of human breast and ovarian cancers have amplified copies of the ERBB2 gene, which codes for a growth factor receptor. The existence of multiple copies of the gene leads to the production of too much receptor protein, which in turn triggers excessive cell proliferation.



Chromosomal translocation: During chromosomal translocation, a portion of one chromosome is physically removed and joined to another chromosome. A classic example occurs in Burkitt's lymphoma, a type of cancer associated with the Epstein-Barr vim (EBV). Infection with EBV stimulates cell proliferation, but it is not sufficient to cause cancer by itself. The disease only arises when a translocation involving chromosome 8 happens to occur in one of these proliferating cells. In the most frequent translocation, a proto-oncogene called MYC is moved from chromosome 8 to 14, where it becomes situated next to an intensely active region of chromosome 14 containing genes coding for antibody molecules Moving the MYC gene so close to the highly active antibody genes causes that gene to also become activated, thereby leading to an overproduction of the Myc protein, a transcription factor that stimulates cell proliferation.

Although the translocated MYC gene retains its normal structure and codes for a normal Myc protein, it is still an oncogene because its new location on chromosome 14 causes the gene to be over-expressed.

Translocations can also disrupt gene structure and cause abnormal proteins to be produced. One example involves the Philadelphia chromosome, an abnormal version of chromosome 22 commonly associated with chronic myelogenous leukaemia. The Philadelphia chromosome is created by DNA breakage near the ends of chromosomes 9 and 22, followed by the reciprocal exchange of DNA between the two chromosomes. This translocation creates an oncogene called BCR-ABL, which contains DNA sequences derived from two different genes (BCR and ABL). As a result, the oncogene produces a fusion protein that functions abnormally because it contains amino acid sequences derived from two different proteins.

Local DNA rearrangements: Another mechanism for creating oncogenes involves local rearrangements in which the base sequences of proto-oncogenes are altered by deletions, insertions, inversions (removal of a sequence followed by reinsertion in the opposite direction), or transpositions (movement of a sequence from one location to another). An example encountered in thyroid and colon cancers illustrates how a simple rearrangement can create an oncogene from two normal genes. This example involves two genes, named NTRK1 and TPM3, that reside on the same chromosome.

NTRK1 codes for a receptor tyrosine kinase and TPM3 codes for a completely unrelated protein, nonmuscle tropomyosin. In some cancers, a DNA inversion occurs that causes one end of the TPM3 gene to fuse to the opposite end of the NTRK1 gene. The resulting gene, called the TRK oncogene, produces a fusion protein containing the tyrosine kinase site of the receptor joined to a region of the tropomyosin molecule that forms coiled coils, which are structures that cause polypeptide chains to join together as dimers. As a result, the fusion protein forms a permanent dimer and its tyrosine kinase is permanently activated. Receptor tyrosine kinases are normally activated by bringing together two receptor molecules to form a dimer.

Insertional mutagenesis: Retroviruses possessing no oncogenes of their own can sometimes cause cancer by integrating their genes into a host chromosome in a region where a host proto-oncogene is located. In such cases, integration of the viral DNA converts the proto-oncogene into an oncogene by either altering the gene's structure and causing it to produce an abnormal protein, or by causing the gene to be overiexpressed. This phenomenon, called insertional mutagenesis, is frequently encountered in animal cancers but is rare in humans.

However, some human cancers have been inadvertently created this way in gene therapy trials that used retroviruses as vectors for repairing defective genes.

The writer is associate professor and head, Department of Botany, Ananda Mohan College, Kolkata