Networks, random walks and statistical optimality

Peter J. Cameron, University of St Andrews

International Webinar on Machine Learning
Central University of Tamil Nadu, Thiruvarur
25 April 2022

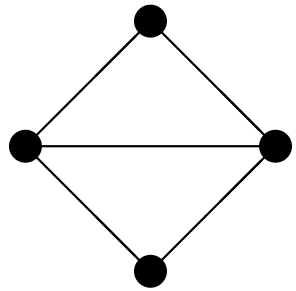
I thank Dr Vijayalakshmi for the invitation to speak to you today.

I thank Dr Vijayalakshmi for the invitation to speak to you today.
I am not an expert on machine learning. But there are strong links between machine learning and networks; and I am going to tell you about networks, and a new technique (whose roots go back to Kirchhoff in the nineteenth century) for studying them, as well as some of their applications in other branches of mathematics and statistics, and elsewhere.

I thank Dr Vijayalakshmi for the invitation to speak to you today.
I am not an expert on machine learning. But there are strong links between machine learning and networks; and I am going to tell you about networks, and a new technique (whose roots go back to Kirchhoff in the nineteenth century) for studying them, as well as some of their applications in other branches of mathematics and statistics, and elsewhere.
From some points of view, a network is the same as a graph, although the different names indicate different emphasis in their study.

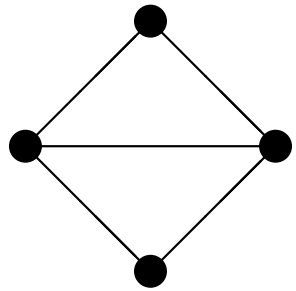
What is a graph?

In its simplest form, a graph has a set of vertices, some pairs of which are joined by edges:

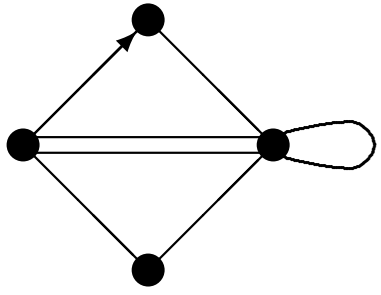


What is a graph?

In its simplest form, a graph has a set of vertices, some pairs of which are joined by edges:

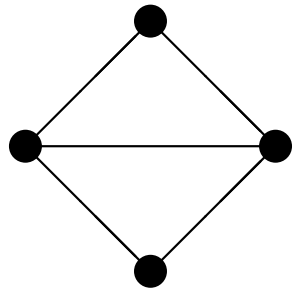


More elaborate graphs could have loops joining vertices to themselves; multiple edges joining the same pairs of vertices, or directed edges like one-way streets:

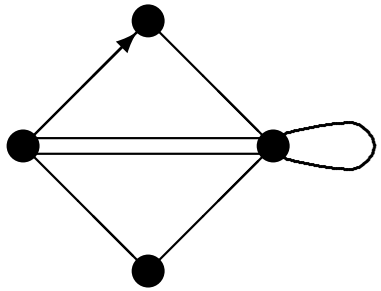


What is a graph?

In its simplest form, a graph has a set of vertices, some pairs of which are joined by edges:



More elaborate graphs could have loops joining vertices to themselves; multiple edges joining the same pairs of vertices, or directed edges like one-way streets:



We can also add further structure such as weights or colours to vertices or edges.

What is a network?

A network is (more or less) the same thing as a graph. The different name indicates that we are thinking about them differently. I will discuss some of the features we commonly meet in networks in the next few slides.

What is a network?

A network is (more or less) the same thing as a graph. The different name indicates that we are thinking about them differently. I will discuss some of the features we commonly meet in networks in the next few slides.

- Networks typically come from some application area.

What is a network?

A network is (more or less) the same thing as a graph. The different name indicates that we are thinking about them differently. I will discuss some of the features we commonly meet in networks in the next few slides.

- Networks typically come from some application area.

Examples include gene networks, where the protein produced by one gene can enhance or inhibit the action of another gene. For example, a simple switch controlled by a single gene on the Y-chromosome determines the sex of a human embryo, but it works by triggering a cascade of gene actions which produce different outcomes.

What is a network?

A network is (more or less) the same thing as a graph. The different name indicates that we are thinking about them differently. I will discuss some of the features we commonly meet in networks in the next few slides.

- Networks typically come from some application area.

Examples include gene networks, where the protein produced by one gene can enhance or inhibit the action of another gene. For example, a simple switch controlled by a single gene on the Y-chromosome determines the sex of a human embryo, but it works by triggering a cascade of gene actions which produce different outcomes.
Another example is in ecology, where the vertices may be species inhabiting the same environment; two species can be in various relations such as predator/prey, symbiotic, parasitic, ...

- Networks are often very large graphs.

- Networks are often very large graphs.

Examples include the internet (nodes are computing devices), the World Wide Web (nodes are web pages), social media, and so on.

- Networks are often very large graphs.

Examples include the internet (nodes are computing devices), the World Wide Web (nodes are web pages), social media, and so on.

- Networks may have some associated dynamics.

- Networks are often very large graphs.

Examples include the internet (nodes are computing devices), the World Wide Web (nodes are web pages), social media, and so on.

- Networks may have some associated dynamics.

This may refer to the network itself (e.g. the Web is constantly growing and changing; what processes are involved?), or to the flow of some commodity such as traffic or information in the network.

- Networks may have some additional structure.

- Networks may have some additional structure.

For example, we saw in the ecological network example that there are several different relations between species. If the relation is predator/prey, then the edge should be directed (say from predator to prey).

- Networks may have some additional structure.

For example, we saw in the ecological network example that there are several different relations between species. If the relation is predator/prey, then the edge should be directed (say from predator to prey).
Similarly the effect of one gene on another can be quite complicated.

- Networks may have some additional structure.

For example, we saw in the ecological network example that there are several different relations between species. If the relation is predator/prey, then the edge should be directed (say from predator to prey).
Similarly the effect of one gene on another can be quite complicated.
But I will give another example, which involves giving some extra structure to the nodes rather than the edges, namely network coding.

Information flow

In the latest supercomputers, each node has processing power, but it cannot export the results of its computations because sending information is much slower than computing it.

Information flow

In the latest supercomputers, each node has processing power, but it cannot export the results of its computations because sending information is much slower than computing it.
So if the computer does a simulation, most of the data it produces is WORN (write once, read never).

Information flow

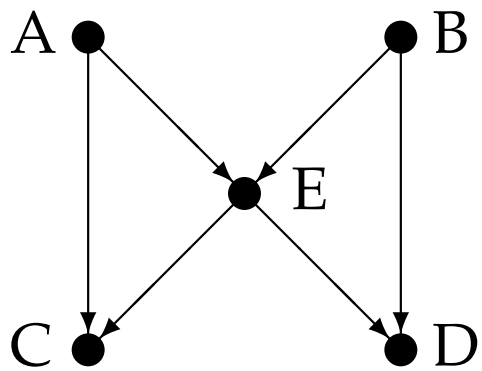
In the latest supercomputers, each node has processing power, but it cannot export the results of its computations because sending information is much slower than computing it.
So if the computer does a simulation, most of the data it produces is WORN (write once, read never).
The usual way to deal with this is to decide which aspects of the data are really important and just export these. But people also look at ways of speeding up data transmission. Network coding is one of these.

Network coding

Network coding can be illustrated with a very simple example, the butterfly network.

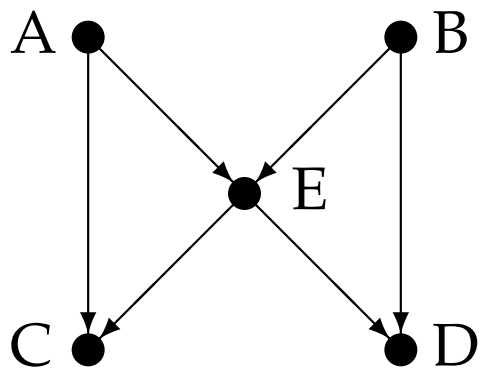
Network coding

Network coding can be illustrated with a very simple example, the butterfly network.



Network coding

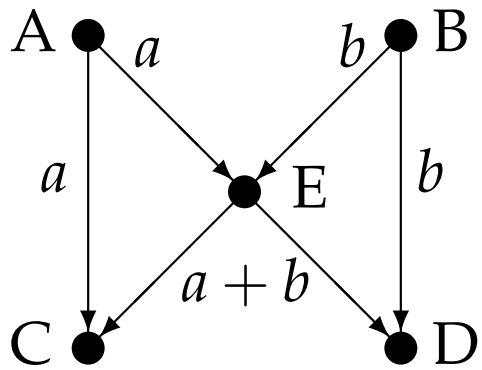
Network coding can be illustrated with a very simple example, the butterfly network.



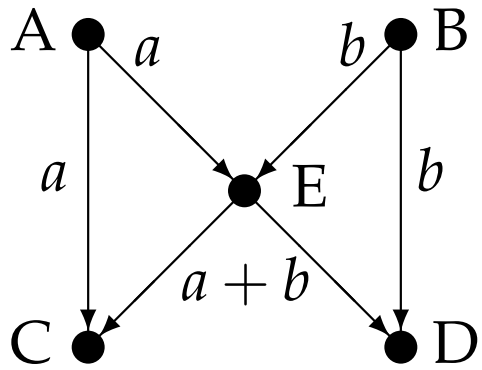
Nodes A and B possess some information in the form of bitstrings a and b. The task is for A to send a to D, and B to send b to C. If it were physical commodities rather than information, it would take two time units to send each bit because of the bottleneck at E.

But if we give each node some small amount of processing power, the information can be sent through the network in just one time unit.

But if we give each node some small amount of processing power, the information can be sent through the network in just one time unit.

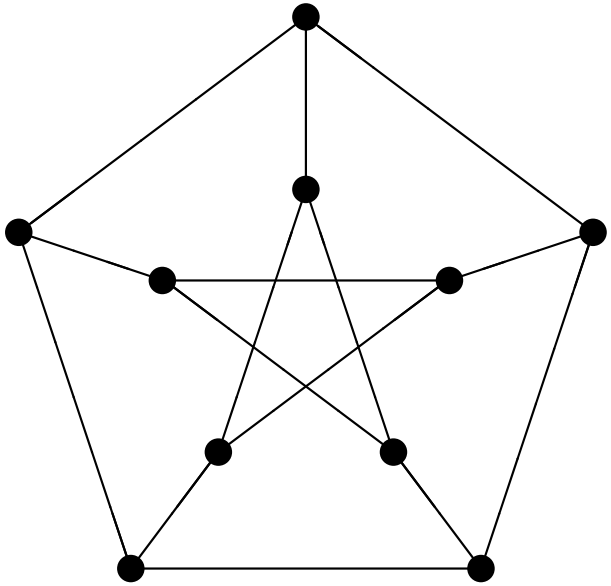


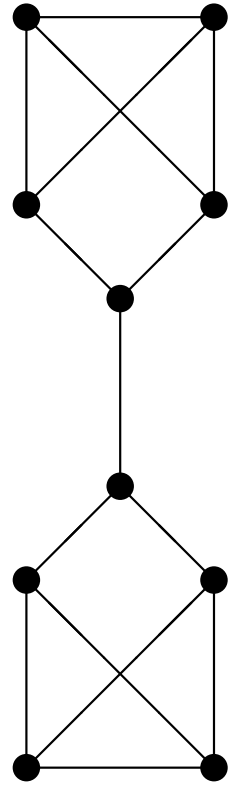
But if we give each node some small amount of processing power, the information can be sent through the network in just one time unit.



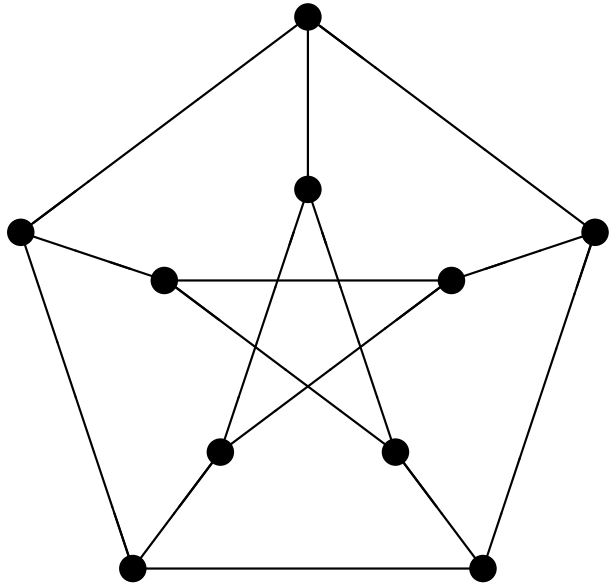
Node A transmits its information a to both C and E, and node B sends b to D and E. Now E adds its two inputs to give $a + b$, which it sends on to C and D. Now C possesses a and $a + b$, and can find b by subtraction; similarly D can find a.

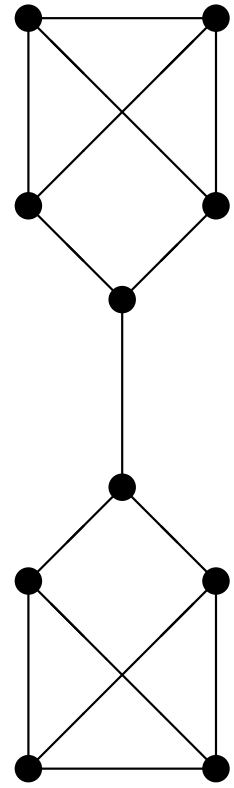
What makes a good network?





What makes a good network?





Suppose you wanted to build a network connecting ten nodes. You could afford to construct fifteen edges. You want the network to be well connected and resilient. Which one of the two shown above would you choose?

I think almost everyone would choose the first network. The second has an obvious bottleneck, a link through which all traffic from top to bottom must pass. This can be quantified in various ways:

I think almost everyone would choose the first network. The second has an obvious bottleneck, a link through which all traffic from top to bottom must pass. This can be quantified in various ways:

- A spanning tree in a graph is a collection of edges forming a connected subgraph with no cycles (a "minimal connector"). The more spanning trees, the better connected. The first graph has 2000 spanning trees, the second 576.

I think almost everyone would choose the first network. The second has an obvious bottleneck, a link through which all traffic from top to bottom must pass. This can be quantified in various ways:

- A spanning tree in a graph is a collection of edges forming a connected subgraph with no cycles (a "minimal connector"). The more spanning trees, the better connected. The first graph has 2000 spanning trees, the second 576.
- Regarding the graph as an electrical network with each edge being a 1-ohm resistor, we can measure the effective resistance between two nodes by connecting a battery and measuring the current that flows. We can sum these over all pairs of vertices. The lower sum the better. For the first graph, the sum is 33; for the second, it is $206/3 = 68.66\ldots$, more than twice as large.

- In a good network there will be many edges out of any set of nodes. The isoperimetric number is the smallest ratio of number of edges from S to its complement to number of vertices in S, for $|S| \leq n/2$. The larger, the better.For the first graph it is 1, and for the second, only 0.2.

- In a good network there will be many edges out of any set of nodes. The isoperimetric number is the smallest ratio of number of edges from S to its complement to number of vertices in S, for $|S| \leq n/2$. The larger, the better.For the first graph it is 1, and for the second, only 0.2.

The intuition is that, if the isoperimetric number is large, you will not get trapped in a small set of vertices (for example if you walk randomly on the graph).

- In a good network there will be many edges out of any set of nodes. The isoperimetric number is the smallest ratio of number of edges from S to its complement to number of vertices in S, for $|S| \leq n/2$. The larger, the better.For the first graph it is 1, and for the second, only 0.2.

The intuition is that, if the isoperimetric number is large, you will not get trapped in a small set of vertices (for example if you walk randomly on the graph).
The isoperimetric number is related to other important properties of a good network such as expansion (where Ramanujan graphs are optimal) and rapid convergence of the random walk on the graph.

All three measures are related to the Laplacian matrix of the network, which I will define shortly. (Note to topologists: networks can be used as discrete approximations to manifolds; the Laplacian matrix is an approximation to the classical Laplacian operator on the manifold.)

All three measures are related to the Laplacian matrix of the network, which I will define shortly. (Note to topologists: networks can be used as discrete approximations to manifolds; the Laplacian matrix is an approximation to the classical Laplacian operator on the manifold.)
The Laplacian matrix and its eigenvalues carry important information about the connectedness of the graph

All three measures are related to the Laplacian matrix of the network, which I will define shortly. (Note to topologists: networks can be used as discrete approximations to manifolds; the Laplacian matrix is an approximation to the classical Laplacian operator on the manifold.)
The Laplacian matrix and its eigenvalues carry important information about the connectedness of the graph
We will also see that the three measures suggested correspond exactly to three popular optimality criteria in experimental design in statistics.

Random walks

The simplest random walk on a graph (which we assume to be connected) is a discrete process which starts at a vertex of the graph. This vertex may be given in advance or selected from some initial probability distribution.

Random walks

The simplest random walk on a graph (which we assume to be connected) is a discrete process which starts at a vertex of the graph. This vertex may be given in advance or selected from some initial probability distribution.
At each time step, the walker moves to a neighbour of the current vertex, all neighbours chosen to be equally likely.

Random walks

The simplest random walk on a graph (which we assume to be connected) is a discrete process which starts at a vertex of the graph. This vertex may be given in advance or selected from some initial probability distribution.
At each time step, the walker moves to a neighbour of the current vertex, all neighbours chosen to be equally likely.
As the number of steps increases, the probability distribution of the position tends to a limiting distribution.

Random walks

The simplest random walk on a graph (which we assume to be connected) is a discrete process which starts at a vertex of the graph. This vertex may be given in advance or selected from some initial probability distribution.
At each time step, the walker moves to a neighbour of the current vertex, all neighbours chosen to be equally likely.
As the number of steps increases, the probability distribution of the position tends to a limiting distribution.
In a more general form, there are positive weights on the edges, and the next step in the random walk is chosen with probability proportional to the weight on the edge.

An example: permutation groups

Here is an example chosen from my own field, permutation groups. We are given a group G which acts on a set X in some way. The set will be partitioned into orbits of the group, an orbit consisting of all points which can be reached from a given one by applying elements of the group.

An example: permutation groups

Here is an example chosen from my own field, permutation groups. We are given a group G which acts on a set X in some way. The set will be partitioned into orbits of the group, an orbit consisting of all points which can be reached from a given one by applying elements of the group.
There may be a large number of orbits, of differing sizes. How do we choose a random orbit (with each orbit, large or small, equally likely to be chosen)?

An example: permutation groups

Here is an example chosen from my own field, permutation groups. We are given a group G which acts on a set X in some way. The set will be partitioned into orbits of the group, an orbit consisting of all points which can be reached from a given one by applying elements of the group.

There may be a large number of orbits, of differing sizes. How do we choose a random orbit (with each orbit, large or small, equally likely to be chosen)?

This was answered by Mark Jerrum, as follows. Form a bipartite graph with $G \cup X$ as vertex set. Put an edge from $g \in G$ to $x \in X$ if the permutation g fixes the element x. Then the limiting distribution of the random walk of even length starting in X is uniform on orbits.

Another example: Latin squares

A Latin square of order n is an $n \times n$ array with entries taken from an alphabet of size n (whose elements may be letters, numbers, colours, etc.) so that each letter occurs once in each row and once in each column.

Another example: Latin squares

A Latin square of order n is an $n \times n$ array with entries taken from an alphabet of size n (whose elements may be letters, numbers, colours, etc.) so that each letter occurs once in each row and once in each column.
Latin squares are used in experimental design in statistics, in cryptography, and in pure mathematics where they occur as "multiplication tables" of algebraic structures (groups, quasigroups, etc.)

Another example: Latin squares

A Latin square of order n is an $n \times n$ array with entries taken from an alphabet of size n (whose elements may be letters, numbers, colours, etc.) so that each letter occurs once in each row and once in each column.

Latin squares are used in experimental design in statistics, in cryptography, and in pure mathematics where they occur as "multiplication tables" of algebraic structures (groups, quasigroups, etc.)

Jacobson and Mathews gave a method for choosing a random Latin square, based in an ingenious way on a set a bit larger than the set of all Latin squares of given order.

Rate of convergence

For many reasons, theoretical and practical, it is important to know how many steps we need to take so that the distribution is within a given distance of its limit. If this takes an exponential number of steps, the algorithm may not be of much use.

Rate of convergence

For many reasons, theoretical and practical, it is important to know how many steps we need to take so that the distribution is within a given distance of its limit. If this takes an exponential number of steps, the algorithm may not be of much use.

This is particularly important in the Jacobson–Mathews random walk for Latin squares; for cryptographic applications we may need to know that the square really is random!

Rate of convergence

For many reasons, theoretical and practical, it is important to know how many steps we need to take so that the distribution is within a given distance of its limit. If this takes an exponential number of steps, the algorithm may not be of much use.

This is particularly important in the Jacobson–Mathews random walk for Latin squares; for cryptographic applications we may need to know that the square really is random!

We will see later the crucial parameter for estimating the rate of convergence of the random walk.

How do we find good networks?

There are results which say that a large random network will have good properties with high probability. But this is maybe not so useful, since designing and analysing a random network may be far from straightforward.

How do we find good networks?

There are results which say that a large random network will have good properties with high probability. But this is maybe not so useful, since designing and analysing a random network may be far from straightforward.
There are two places we might look.

How do we find good networks?

There are results which say that a large random network will have good properties with high probability. But this is maybe not so useful, since designing and analysing a random network may be far from straightforward.
There are two places we might look.

- Naturally occurring networks such has gene networks, which have been refined by the pressure of millions of years of evolution, are likely to have good properties. But of course, the "selfish gene" might have a view of what makes a good network which is rather different from ours.

- The other place we can look is algebra, specifically group theory. Groups were invented to measure symmetry, but they can be used to construct networks which have good properties, and also provide techniques to analyse and quantify these properties, based on the symmetry of the networks.

- The other place we can look is algebra, specifically group theory. Groups were invented to measure symmetry, but they can be used to construct networks which have good properties, and also provide techniques to analyse and quantify these properties, based on the symmetry of the networks.

The basic construction goes back to Cayley in the 19th century; the objects are known as Cayley graphs. Now representation theory, which grew out of group theory, allows us to estimate the eigenvalues of large networks, which (as we will see) gives strong information about their properties.

- The other place we can look is algebra, specifically group theory. Groups were invented to measure symmetry, but they can be used to construct networks which have good properties, and also provide techniques to analyse and quantify these properties, based on the symmetry of the networks.

The basic construction goes back to Cayley in the 19th century; the objects are known as Cayley graphs. Now representation theory, which grew out of group theory, allows us to estimate the eigenvalues of large networks, which (as we will see) gives strong information about their properties.
Before looking at this, I will describe one more network which has captured the interest of mathematicians.

Erdős number

What about the network in which the nodes are mathematicians, two nodes being joined by an edge if the two mathematicians have published a joint paper.

Erdős number

What about the network in which the nodes are mathematicians, two nodes being joined by an edge if the two mathematicians have published a joint paper.
The distance in this network is called collaboration distance, and a mathematician's distance from Paul Erdős is her Erdős number.

Erdős number

What about the network in which the nodes are mathematicians, two nodes being joined by an edge if the two mathematicians have published a joint paper.
The distance in this network is called collaboration distance, and a mathematician's distance from Paul Erdős is her Erdős number.
The MathSciNet website includes a collaboration distance calculator.

Erdős number

What about the network in which the nodes are mathematicians, two nodes being joined by an edge if the two mathematicians have published a joint paper.
The distance in this network is called collaboration distance, and a mathematician's distance from Paul Erdős is her Erdős number.
The MathSciNet website includes a collaboration distance calculator.
Jerry Grossman at Oakland University in the USA keeps the data for this. There are roughly 400000 authors of mathematics papers on MathSciNet (so the network has 400000 nodes). Of these, about 268000 lie in a single large connected component; 84000 are isolated nodes (mathematicians who have never written a joint paper), amd the remainder, about 50000, lie in small components with between 2 and 23 vertices. (These figures date from 2004 so the picture is probably different now.)

Variants

G. H. Hardy and J. E. Littlewood had a very successful collaboration, and wrote over 100 joint papers, many of them of groundbreaking importance in a number of areas of mathematics.

Variants

G. H. Hardy and J. E. Littlewood had a very successful collaboration, and wrote over 100 joint papers, many of them of groundbreaking importance in a number of areas of mathematics.
As a result, some people have suggested that the distance between Hardy and Littlewood should be smaller, say 1/100.

Variants

G. H. Hardy and J. E. Littlewood had a very successful collaboration, and wrote over 100 joint papers, many of them of groundbreaking importance in a number of areas of mathematics.
As a result, some people have suggested that the distance between Hardy and Littlewood should be smaller, say 1/100. Now Hardy had 8 joint papers with Srinivasa Ramanujan, whereas Littlewood did not have any. So what should the distance from Littlewood to Ramanujan be? Should it be smaller because of their common co-author?

Variants

G. H. Hardy and J. E. Littlewood had a very successful collaboration, and wrote over 100 joint papers, many of them of groundbreaking importance in a number of areas of mathematics.
As a result, some people have suggested that the distance between Hardy and Littlewood should be smaller, say 1/100. Now Hardy had 8 joint papers with Srinivasa Ramanujan, whereas Littlewood did not have any. So what should the distance from Littlewood to Ramanujan be? Should it be smaller because of their common co-author?
I am going to propose a possible answer to this question in the rest of this lecture.

Resistance distance

Take a network, and suppose (to start) that the vertices are terminals in an electrical network, and each edge is a 1-ohm resistor. As we saw earlier, we can measure the effective resistance between any pair of terminals by connecting a battery to those terminals and measuring the current that flows; or we can calculate the effective resistance using Kirchhoff's and Ohm's Laws.

Resistance distance

Take a network, and suppose (to start) that the vertices are terminals in an electrical network, and each edge is a 1-ohm resistor. As we saw earlier, we can measure the effective resistance between any pair of terminals by connecting a battery to those terminals and measuring the current that flows; or we can calculate the effective resistance using Kirchhoff's and Ohm's Laws.

More generally, we could connect arbitrary resistors between the terminals, and measure effective resistance betweem each pair of terminals.

Resistance distance

Take a network, and suppose (to start) that the vertices are terminals in an electrical network, and each edge is a 1-ohm resistor. As we saw earlier, we can measure the effective resistance between any pair of terminals by connecting a battery to those terminals and measuring the current that flows; or we can calculate the effective resistance using Kirchhoff's and Ohm's Laws.

More generally, we could connect arbitrary resistors between the terminals, and measure effective resistance betweem each pair of terminals.

Something remarkable happens:

Theorem

Effective resistance is a metric on the network.

Resistance distance

Take a network, and suppose (to start) that the vertices are terminals in an electrical network, and each edge is a 1-ohm resistor. As we saw earlier, we can measure the effective resistance between any pair of terminals by connecting a battery to those terminals and measuring the current that flows; or we can calculate the effective resistance using Kirchhoff's and Ohm's Laws.

More generally, we could connect arbitrary resistors between the terminals, and measure effective resistance betweem each pair of terminals.

Something remarkable happens:

Theorem

Effective resistance is a metric on the network.

This means that it is positive, symmetric (doesn't depend on the order of the terminals), and satisfies the triangle inequality:

$$R(i,j) + R(j,k) \geq R(i,k).$$

I will not prove this theorem in detail. It depends on a fact which is obvious to any physicist or electrical engineer, but is not trivial for a mathematician to prove:

I will not prove this theorem in detail. It depends on a fact which is obvious to any physicist or electrical engineer, but is not trivial for a mathematician to prove:

> *If we connect a battery between two terminals of a network, the potential at any other terminal will be between the potentials of the two terminals connected to the battery.*

I will not prove this theorem in detail. It depends on a fact which is obvious to any physicist or electrical engineer, but is not trivial for a mathematician to prove:

> *If we connect a battery between two terminals of a network, the potential at any other terminal will be between the potentials of the two terminals connected to the battery.*

The other fact needed in the proof is that the equations for the network (Kirchhoff's and Ohm's laws) are linear, so if we have two solutions, we can superimpose them to get another solution by adding the currents.

Conductance

It is convenient to use conductance instead of resistance for the network elements. Conductance is just the reciprocal of resistance; its units are mhos. By the law for parallel connection, if several edges connect the same pair of terminals, we can replace them by a single edge by adding the conductances.

Conductance

It is convenient to use conductance instead of resistance for the network elements. Conductance is just the reciprocal of resistance; its units are mhos. By the law for parallel connection, if several edges connect the same pair of terminals, we can replace them by a single edge by adding the conductances.
This agrees with the earlier intuition that, if Hardy and Littlewood wrote 100 joint papers, then the conductance of the edge joining them should be 100.

Conductance

It is convenient to use conductance instead of resistance for the network elements. Conductance is just the reciprocal of resistance; its units are mhos. By the law for parallel connection, if several edges connect the same pair of terminals, we can replace them by a single edge by adding the conductances.
This agrees with the earlier intuition that, if Hardy and Littlewood wrote 100 joint papers, then the conductance of the edge joining them should be 100.
Another advantage is that, if two nodes are not joined by an edge, then the resistance between them should be infinite, in other words, the conductance should be 0.

Conductance

It is convenient to use conductance instead of resistance for the network elements. Conductance is just the reciprocal of resistance; its units are mhos. By the law for parallel connection, if several edges connect the same pair of terminals, we can replace them by a single edge by adding the conductances.
This agrees with the earlier intuition that, if Hardy and Littlewood wrote 100 joint papers, then the conductance of the edge joining them should be 100.
Another advantage is that, if two nodes are not joined by an edge, then the resistance between them should be infinite, in other words, the conductance should be 0.
Then resistance distance would be a very good measure of collaboration distance in the collaboration network.

Conductance

It is convenient to use conductance instead of resistance for the network elements. Conductance is just the reciprocal of resistance; its units are mhos. By the law for parallel connection, if several edges connect the same pair of terminals, we can replace them by a single edge by adding the conductances.
This agrees with the earlier intuition that, if Hardy and Littlewood wrote 100 joint papers, then the conductance of the edge joining them should be 100.
Another advantage is that, if two nodes are not joined by an edge, then the resistance between them should be infinite, in other words, the conductance should be 0.
Then resistance distance would be a very good measure of collaboration distance in the collaboration network.
As far as I know, nobody has ever computed resistance distances in this network.

Laplacian matrix

Here is a method for computing resistance distance. It is essentially due to Kirchhoff in the 19th century.

Laplacian matrix

Here is a method for computing resistance distance. It is essentially due to Kirchhoff in the 19th century.
We are given a network with n nodes; let s_{ij} be the conductance of the direct connection between nodes i and j. So $s_{ij} = 0$ if there is no edge between these nodes.

Laplacian matrix

Here is a method for computing resistance distance. It is essentially due to Kirchhoff in the 19th century.
We are given a network with n nodes; let s_{ij} be the conductance of the direct connection between nodes i and j. So $s_{ij} = 0$ if there is no edge between these nodes.
Form the Laplacian matrix L of the network, an $n \times n$ matrix with rows and columns indexed by network nodes, as follows. For $i \neq j$, the (i, j) entry is $-s_{ij}$, while the (i, i) diagonal entry is the sum of all conductances between i and other vertices.

Laplacian matrix

Here is a method for computing resistance distance. It is essentially due to Kirchhoff in the 19th century.
We are given a network with n nodes; let s_{ij} be the conductance of the direct connection between nodes i and j. So $s_{ij} = 0$ if there is no edge between these nodes.
Form the Laplacian matrix L of the network, an $n \times n$ matrix with rows and columns indexed by network nodes, as follows. For $i \neq j$, the (i, j) entry is $-s_{ij}$, while the (i, i) diagonal entry is the sum of all conductances between i and other vertices.
This matrix has all row and column sums zero; such a matrix has the property that all order $n - 1$ cofactors (determinants obtained by deleting the ith row and jth column, with a change of sign if $i + j$ is odd) are equal. This common value A counts the spanning trees in the network, if all conductances are 0 or 1.
In general, it counts spanning trees by weight, where the weight of a tree is the product of the conductances of its edges.

Connected components and random walks

Since the row and column sumns of L are all zero, we know that L is singular; indeed, the all-1 vector is an eigenvector with eigenvalue 0 (called the trivial eigenvalue and eigenvector).

Connected components and random walks

Since the row and column sumns of L are all zero, we know that L is singular; indeed, the all-1 vector is an eigenvector with eigenvalue 0 (called the trivial eigenvalue and eigenvector). It can be shown that the multiplicity of 0 as an eigenvalue is equal to the number of connected components of the graph.

Connected components and random walks

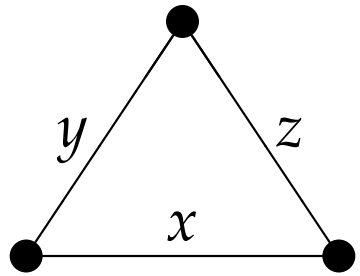
Since the row and column sumns of L are all zero, we know that L is singular; indeed, the all-1 vector is an eigenvector with eigenvalue 0 (called the trivial eigenvalue and eigenvector). It can be shown that the multiplicity of 0 as an eigenvalue is equal to the number of connected components of the graph. Indeed, the size of the smallest non-trivial eigenvalue is an important measure of the connectivity of the graph.

Connected components and random walks

Since the row and column sumns of L are all zero, we know that L is singular; indeed, the all-1 vector is an eigenvector with eigenvalue 0 (called the trivial eigenvalue and eigenvector).
It can be shown that the multiplicity of 0 as an eigenvalue is equal to the number of connected components of the graph.
Indeed, the size of the smallest non-trivial eigenvalue is an important measure of the connectivity of the graph.
This shows up in various ways. For a regular graph of valency k, the smallest non-trivial eigenvalue is at most $k - 2\sqrt{k-1}$; equality implies that the graph is a Ramanujan graph. In general, a large value of the smallest eigenvalue is associated with a large isoperimetric number, and with rapid convergence of the random walk on the graph.

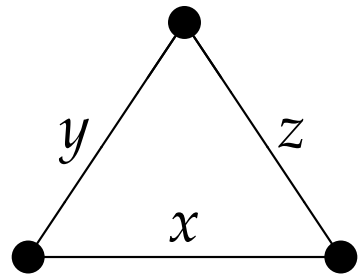
Furthermore, if we delete the ith and jth row and the ith and jth column, then the value is not independent of i and j; it counts spanning 2-forests (forests with two connected components) in which the nodes i and j lie in different components. If this value is B_{ij}, then the effective resistance between i and j is B_{ij}/A.

Furthermore, if we delete the ith and jth row and the ith and jth column, then the value is not independent of i and j; it counts spanning 2-forests (forests with two connected components) in which the nodes i and j lie in different components. If this value is B_{ij}, then the effective resistance between i and j is B_{ij}/A.



For example, suppose we have a triangle with conductances x, y and z. Then the effective resistance between the ends of the edge with conductance x is $(y+z)/(xy+yz+zx)$.

Furthermore, if we delete the ith and jth row and the ith and jth column, then the value is not independent of i and j; it counts spanning 2-forests (forests with two connected components) in which the nodes i and j lie in different components. If this value is B_{ij}, then the effective resistance between i and j is B_{ij}/A.



For example, suppose we have a triangle with conductances x, y and z. Then the effective resistance between the ends of the edge with conductance x is $(y+z)/(xy+yz+zx)$.
I would be very interested in a calculation of the resistance distances in the mathematicians' collaboration graph!

Statistics

I spoke earlier about three ways of deciding which network is "good", according to three different criteria. These correspond precisely to three optimality conditions in experimental design in statistics, which I now briefly mention.

Statistics

I spoke earlier about three ways of deciding which network is "good", according to three different criteria. These correspond precisely to three optimality conditions in experimental design in statistics, which I now briefly mention.
The set-up is very general: the treatments might be fertilizer regimes for crops, and the experimental units plots of land; the treatments may be drugs, and the experimental units patients; and so on.

Statistics

I spoke earlier about three ways of deciding which network is "good", according to three different criteria. These correspond precisely to three optimality conditions in experimental design in statistics, which I now briefly mention.

The set-up is very general: the treatments might be fertilizer regimes for crops, and the experimental units plots of land; the treatments may be drugs, and the experimental units patients; and so on.

The situation is that we have a number n of "treatments" to compare, and a number of "experimental units" to apply them to. What is the best way to do this? For the most accurate experimental result, we want to minimize the variances between the estimators of treatment differences; but this is a multidimensional problem, so we cannot minimize all variances simultaneously!

If all experimental units are exactly alike, then we simply apply each treatment to the same number of experimental units (as near as possible). But a more usual situation is that the experimental units are grouped into blocks, so that units in a block are alike but differ systematically from those in a different block.

If all experimental units are exactly alike, then we simply apply each treatment to the same number of experimental units (as near as possible). But a more usual situation is that the experimental units are grouped into blocks, so that units in a block are alike but differ systematically from those in a different block.

For example, in the fertilizer example, blocks might be different farms or areas of the country; in the drug example, they might be hospitals in which the tests are administered.

If all experimental units are exactly alike, then we simply apply each treatment to the same number of experimental units (as near as possible). But a more usual situation is that the experimental units are grouped into blocks, so that units in a block are alike but differ systematically from those in a different block.

For example, in the fertilizer example, blocks might be different farms or areas of the country; in the drug example, they might be hospitals in which the tests are administered.

In this situation, we form the concurrence graph of the design, where the vertices are the treatments, and we put an edge between treatments i and j each time there is a block containing both. So the graph can have multiple edges. It is important to realise that the graph can be chosen by the experimenter as part of the design, subject to constraints about the numbers and sizes of blocks.

Optimal designs

The effective resistance between i and j is proportional to the variance of the estimator of the difference in effect between these two treatments. So we want to keep these variances small.

Optimal designs

The effective resistance between i and j is proportional to the variance of the estimator of the difference in effect between these two treatments. So we want to keep these variances small.

- If we minimize the average resistance, the average variance is minimized: such a design is called A-optimal (A for "average").

Optimal designs

The effective resistance between i and j is proportional to the variance of the estimator of the difference in effect between these two treatments. So we want to keep these variances small.

- If we minimize the average resistance, the average variance is minimized: such a design is called A-optimal (A for "average").
- If we minimize the largest variance, and hence the largest resistance (this is roughly the same as maximizing the isoperimetric number), the design is E-optimal (E for "extreme").

Optimal designs

The effective resistance between i and j is proportional to the variance of the estimator of the difference in effect between these two treatments. So we want to keep these variances small.

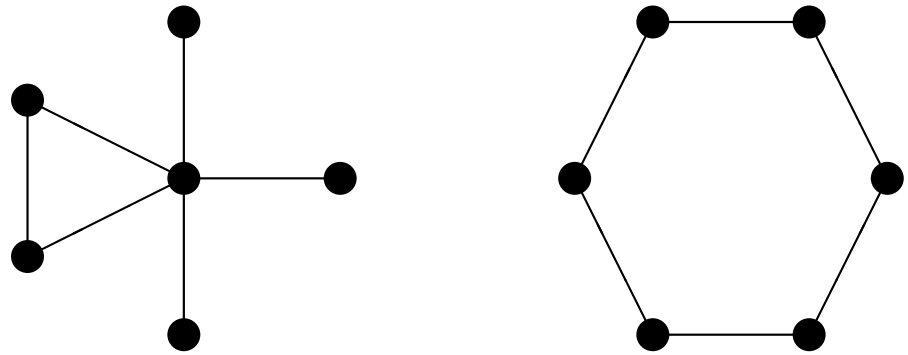
- If we minimize the average resistance, the average variance is minimized: such a design is called A-optimal (A for "average").
- If we minimize the largest variance, and hence the largest resistance (this is roughly the same as maximizing the isoperimetric number), the design is E-optimal (E for "extreme").
- Minimizing the volume of a confidence ellipsoid containing the estimated values is equivalent to maximizing the count of spanning trees by weight; such a design is D-optimal (D for "determinant").

An example

Suppose that we have n treatments to compare, and $2n$ experimental units divided into n blocks of size 2. What is the best design?

An example

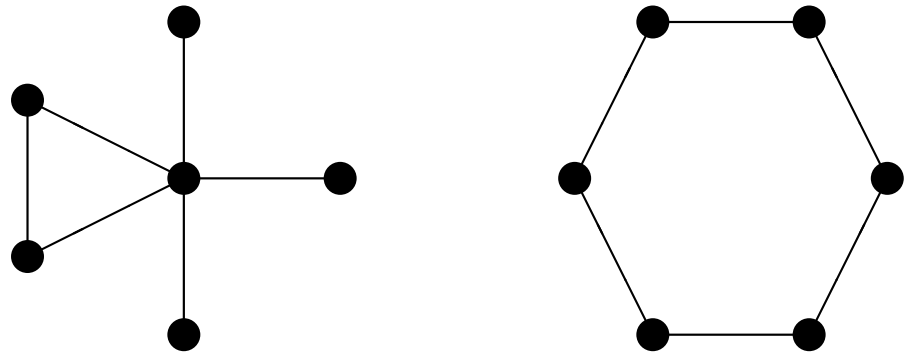
Suppose that we have n treatments to compare, and $2n$ experimental units divided into n blocks of size 2. What is the best design?



Biologists are told that every treatment should be compared with a control, so would use the design on the left. Statisticians are told that each treatment should be replicated equally often, so would use the design on the right.

An example

Suppose that we have n treatments to compare, and $2n$ experimental units divided into n blocks of size 2. What is the best design?



Biologists are told that every treatment should be compared with a control, so would use the design on the left. Statisticians are told that each treatment should be replicated equally often, so would use the design on the right.

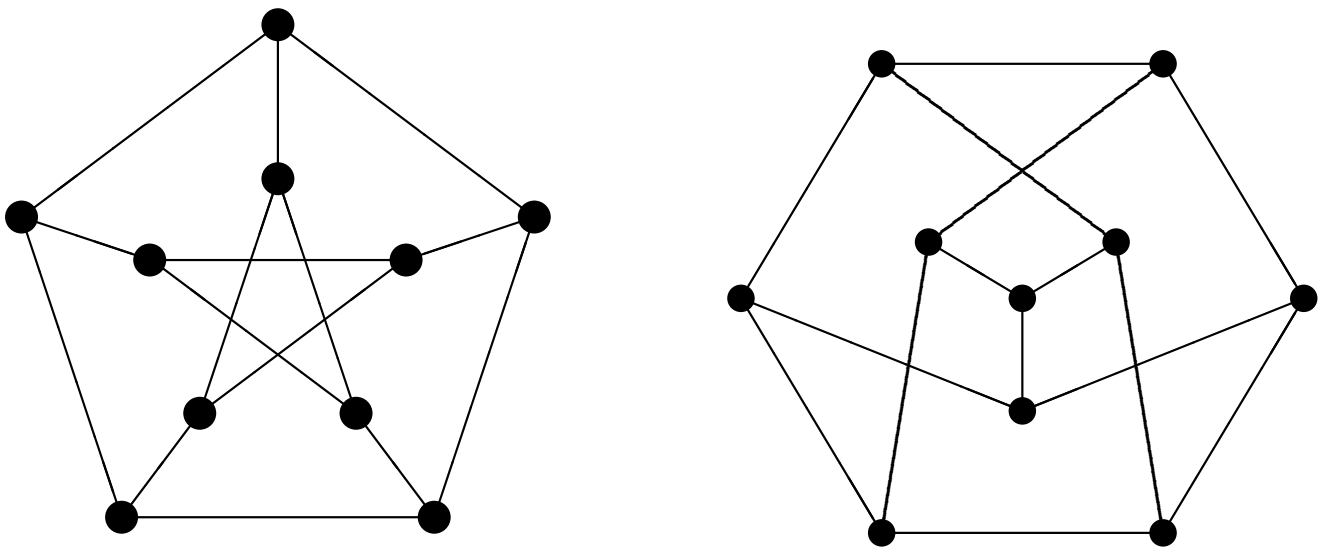
The cycle has n spanning trees, the design on the left just three, so the cycle wins on the D-criterion. However, it can be shown that the design on the left (the "queen bee design") is A-optimal if $n \geq 12$ and E-optimal if $n \geq 7$, a result which came as a surprise to some statisticians.

Graph isomorphism

Two graphs or networks are isomorphic if the vertices of the first can be matched up with those of the second so that corresponding pairs of vertices are either both joined or both not joined (and if the edges are weighted, these edges should have the same weight).

Graph isomorphism

Two graphs or networks are isomorphic if the vertices of the first can be matched up with those of the second so that corresponding pairs of vertices are either both joined or both not joined (and if the edges are weighted, these edges should have the same weight).
For example, these two graphs are isomorphic:



The Graph Isomorphism Problem

Consider the following computational problem:

> *Given two graphs, decide whether or not they are isomorphic.*

The Graph Isomorphism Problem

Consider the following computational problem:

> *Given two graphs, decide whether or not they are isomorphic.*

This is one of a very small group of natural problems which belong to the complexity class NP (this means that we can check whether a proposed isomorphism really works) but are not known to be either in P or NP-complete.

The Graph Isomorphism Problem

Consider the following computational problem:

> *Given two graphs, decide whether or not they are isomorphic.*

This is one of a very small group of natural problems which belong to the complexity class NP (this means that we can check whether a proposed isomorphism really works) but are not known to be either in P or NP-complete.
It is easy to find negative tests. For example, if the graphs have different numbers of vertices or of edges, or if the lists of vertex degrees are different, then they cannot be isomorphic.

One of the first serious attacks on this problem was the Weisfeiler–Leman algorithm. It is based on a refinement of the ideas above, refining the graph by adding purely combinatorial labels to the vertices and edges to distinguish them. The end result is an object known as a coherent configuration; if the two coherent configurations are not isomorphic then the graphs are not isomorphic.

One of the first serious attacks on this problem was the Weisfeiler–Leman algorithm. It is based on a refinement of the ideas above, refining the graph by adding purely combinatorial labels to the vertices and edges to distinguish them. The end result is an object known as a coherent configuration; if the two coherent configurations are not isomorphic then the graphs are not isomorphic.

The algorithm does not always succeed; but using these ideas, László Babai has recently given a quasi-polynomial algorithm for graph isomorphism (running in time $\exp(O((\log n)^c))$ for some c (here $c = 1$ would be polynomial-time).

The RDT algorithm

The WL algorithm may take many steps to reach the coherent configuration. Since it only explores the graph locally, the number of steps is at least $\log d$, where d is the diameter of the graph.

The RDT algorithm

The WL algorithm may take many steps to reach the coherent configuration. Since it only explores the graph locally, the number of steps is at least $\log d$, where d is the diameter of the graph.

Michael Kagan has proposed an alternative approach based on resistance distance, as follows.

The RDT algorithm

The WL algorithm may take many steps to reach the coherent configuration. Since it only explores the graph locally, the number of steps is at least $\log d$, where d is the diameter of the graph.
Michael Kagan has proposed an alternative approach based on resistance distance, as follows.

- Calculate the resistance distances between all pairs of vertices.

The RDT algorithm

The WL algorithm may take many steps to reach the coherent configuration. Since it only explores the graph locally, the number of steps is at least $\log d$, where d is the diameter of the graph.
Michael Kagan has proposed an alternative approach based on resistance distance, as follows.

- Calculate the resistance distances between all pairs of vertices.
- Now give each pair of vertices a link whose conductance is the reciprocal of the effective distance between them.

The RDT algorithm

The WL algorithm may take many steps to reach the coherent configuration. Since it only explores the graph locally, the number of steps is at least $\log d$, where d is the diameter of the graph.
Michael Kagan has proposed an alternative approach based on resistance distance, as follows.

- Calculate the resistance distances between all pairs of vertices.
- Now give each pair of vertices a link whose conductance is the reciprocal of the effective distance between them.
- Repeat until the process stabilises.

The RDT algorithm

The WL algorithm may take many steps to reach the coherent configuration. Since it only explores the graph locally, the number of steps is at least $\log d$, where d is the diameter of the graph.
Michael Kagan has proposed an alternative approach based on resistance distance, as follows.

- Calculate the resistance distances between all pairs of vertices.
- Now give each pair of vertices a link whose conductance is the reciprocal of the effective distance between them.
- Repeat until the process stabilises.

This is currently being investigated. Because it is global, it usually takes many fewer steps than the WL algorithm. We think that the stable configurations are Jordan schemes, related to coherent configurations as Jordan algebras are to associative algebras.

Jordan algebras

Pascual Jordan was an early pioneer in quantum mechanics, a student of Max Born and colleague of Werner Heisenberg. He proposed his algebras for their potential use in quantum mechanics, and was involved in the Jordan–von Neumann–Wigner theorem classifying the simple ones (the analogue of Wedderburn's theorem).

Jordan algebras

Pascual Jordan was an early pioneer in quantum mechanics, a student of Max Born and colleague of Werner Heisenberg. He proposed his algebras for their potential use in quantum mechanics, and was involved in the Jordan–von Neumann–Wigner theorem classifying the simple ones (the analogue of Wedderburn's theorem).
Rather than give you the axioms, I will simply describe an important class of examples. Define a multiplication on the set of $n \times n$ real symmetric matrices by the rule

$$A \circ B = \tfrac{1}{2}(AB + BA).$$

The operation is commutative but not associative.

Jordan algebras

Pascual Jordan was an early pioneer in quantum mechanics, a student of Max Born and colleague of Werner Heisenberg. He proposed his algebras for their potential use in quantum mechanics, and was involved in the Jordan–von Neumann–Wigner theorem classifying the simple ones (the analogue of Wedderburn's theorem).
Rather than give you the axioms, I will simply describe an important class of examples. Define a multiplication on the set of $n \times n$ real symmetric matrices by the rule

$$A \circ B = \tfrac{1}{2}(AB + BA).$$

The operation is commutative but not associative.
Jordan algebras are not much used in physics now but have found a place in statistics and in algebraic combinatorics.

. . . for your attention.