MATHEMATICAL FRONTIERS

The National Academies of SCIENCES ENGINEERING MEDICINE

nas.edu/MathFrontiers

Board on
Mathematical Sciences & Analytics

MATHEMATICAL FRONTIERS
2018 Monthly Webinar Series, 2-3pm ET

February 13*:
Mathematics of the Electric Grid

March 13*:
Probability for People and Places

April 10*:
Social and Biological Networks

May 8*:
Mathematics of Redistricting

June 12*: *Number Theory: The Riemann Hypothesis*

July 10*: *Topology*

August 14*: *Algorithms for Threat Detection*

September 11*: *Mathematical Analysis*

October 9: *Combinatorics*

November 13:
Why Machine Learning Works

December 11:
Mathematics of Epidemics

**** Recording posted***

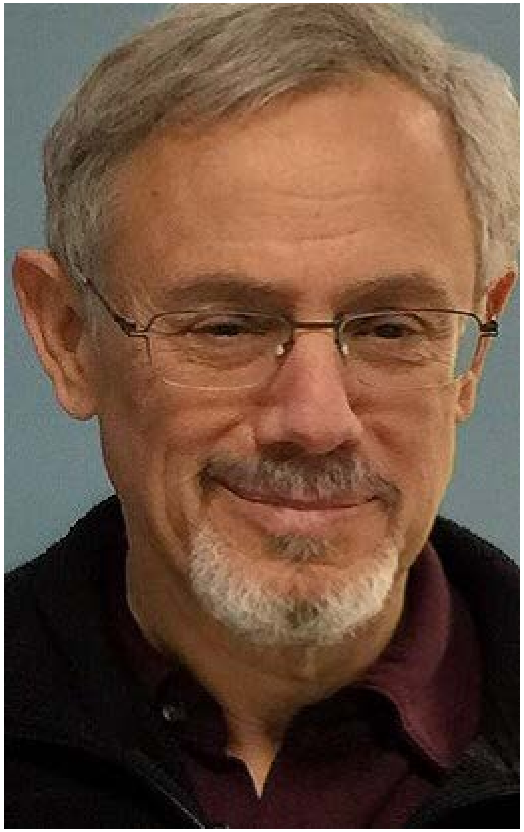
Made possible by support for BMSA from the National Science Foundation Division of Mathematical Sciences and the Department of Energy Advanced Scientific Computing Research

View webinar videos and learn more about BMSA at www.nas.edu/MathFrontiers

MATHEMATICAL FRONTIERS
Combinatorics

Sara Billey,
University of Washington

Jacques Verstraete,
University of California, San Diego



Mark Green,
UCLA (moderator)

View webinar videos and learn more about BMSA at www.nas.edu/MathFrontiers

MATHEMATICAL FRONTIERS
Combinatorics

Sara Billey,
University of Washington

John Rainwater Faculty Fellow and Professor of Mathematics in the Department of Mathematics at the University of Washington

What is Combinatorics?

What is Combinatorics?

Combinatorics is

the nanotechnology of mathematics

This technology applies to problems on

- Existence
- Enumeration
- Optimization

of discrete structures taking into account constraints, patterns, preferences, and rules.

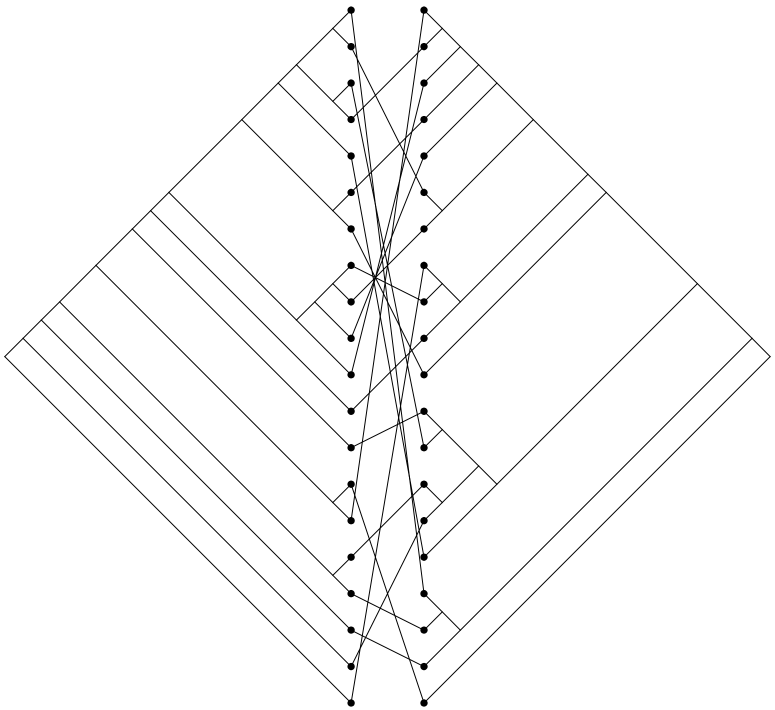
Applications

In the past 100 years, combinatorics has revolutionized the way we think about problems in

- Biology
- Chemistry
- Computer Science
- Physics
- Industry
- Government
- Mathematics

Examples

- The Stable Matching Algorithm
- Tanglegrams



Example 1: Stable Matching

- In 1952, the **National Resident Matching Program (NRMP)** introduced an algorithm to match medical students to residency positions at hospitals in a way that respects the preferences of the students and hospitals without any there being any student-hospital pair who prefer each other over their assignment.

- In 1962, David Gale and Lloyd Shapley proved that the algorithm always produces an assignment which is simultaneously optimal for all students among all stable matchings.

- In 2012, Lloyd Shapley and Alvin Roth won the Nobel prize in Economics for their work realizing other non-monetary markets where the Stable Match Algorithm should be applied: kidney donation.

- How does it work?

Stable Matching: How does it work?

- Students and hospitals each input a ranked list showing their preferences for the match.

Student Prfs	Boston	Houston	Seattle
Andrea	2	1	3
Lakshmi	1	3	2
Ming	2	1	3

Hospital Prfs	Andrea	Lakshmi	Ming
Boston	2	3	1
Houston	2	3	1
Seattle	3	2	1

Stable Matching: How does it work?

- Students and hospitals each input a ranked list showing their preferences for the match.

Student Prfs	Boston	Houston	Seattle
Andrea	2	1	3
Lakshmi	1	3	2
Ming	2	1	3

Hospital Prfs	Andrea	Lakshmi	Ming
Boston	2	3	1
Houston	2	3	1
Seattle	3	2	1

Stable Matching: How does it work?

- Students and hospitals each input a ranked list showing their preferences for the match.

Student Prfs	Boston	Houston	Seattle
Andrea	2	1	3
Lakshmi	1	3	2
Ming	2	1	3

Hospital Prfs	Andrea	Lakshmi	Ming
Boston	2	3	1
Houston	2	3	1
Seattle	3	2	1

Stable Matching: How does it work?

- Students and hospitals each input a ranked list showing their preferences for the match.

Student Prfs	Boston	Houston	Seattle
Andrea	2	1	3
Lakshmi	1	3	2
Ming	2	1	3

One match:
Andrea – Houston,
Lakshmi – Boston,
Ming –Seattle

Hospital Prfs	Andrea	Lakshmi	Ming
Boston	2	3	1
Houston	2	3	1
Seattle	3	2	1

Unstable pair:
Ming – Houston
prefer each other
over their
assignment

Key Questions

- **Definition**: An assignment of students to hospitals is a stable matching if no student and hospital prefer each other over the one given by the assignment.
- **Existence Question**: Given any input preferences of n students and n hospitals, does a stable matching always exist?
- **Enumeration Question**: If so, how many stable matchings are there at most for n students and n hospitals?
- **Optimization Question**: Given any input preferences of n students and n hospitals, what is the best possible assignment for students? For hospitals?

Stable Match Algorithm

- Each student "proposes" to their first choice of hospital.
- Then, hospitals reject all but their highest ranked proposal.
- Rinse, lather, repeat!

Student Prfs	Boston	Houston	Seattle
Andrea	2	1	3
Lakshmi	1	3	2
Ming	2	1	3

Hospital Prfs	Andrea	Lakshmi	Ming
Boston	2	3	1
Houston	2	3	1
Seattle	3	2	1

Stable Match Algorithm

- Each student “proposes” to their first choice of hospital.
- Then, hospitals reject all but their highest ranked proposal.
- Rinse, lather, repeat!

Student Prfs	Boston	Houston	Seattle
Andrea	2	~~1~~	3
Lakshmi	1	3	2
Ming	2	1	3

Hospital Prfs	Andrea	Lakshmi	Ming
Boston	2	3	1
Houston	2	3	1
Seattle	3	2	1

Stable Match Algorithm

- Each student "proposes" to their first choice of hospital.
- Then, hospitals reject all but their highest ranked proposal.
- Rinse, lather, repeat!

Student Prfs	Boston	Houston	Seattle
Andrea	2	~~1~~	3
Lakshmi	1	3	2
Ming	2	1	3

Hospital Prfs	Andrea	Lakshmi	Ming
Boston	2	3	1
Houston	2	3	1
Seattle	3	2	1

Stable Match Algorithm

- Each student "proposes" to their first choice of hospital.
- Then, hospitals reject all but their highest ranked proposal.
- Rinse, lather, repeat!

Student Prfs	Boston	Houston	Seattle
Andrea	2	~~1~~	3
Lakshmi	~~1~~	3	2
Ming	2	1	3

Hospital Prfs	Andrea	Lakshmi	Ming
Boston	2	3	1
Houston	2	3	1
Seattle	3	2	1

Stable Match Algorithm

- Each student "proposes" to their first choice of hospital.
- Then, hospitals reject all but their highest ranked proposal.
- Rinse, lather, repeat!

Student Prfs	Boston	Houston	Seattle
Andrea	2	~~1~~	3
Lakshmi	~~1~~	3	2
Ming	2	1	3

Another match:
Andrea – Boston,
Lakshmi – Seattle,
Ming – Houston

Hospital Prfs	Andrea	Lakshmi	Ming
Boston	2	3	1
Houston	2	3	1
Seattle	3	2	1

Stable!
No pair wants to disregard this assignment.

Stable Match Algorithm

- **Theorem** (Gale-Shapley): For any input preferences by n students and n hospitals, the Stable Match Algorithm produces an assignment with no unstable pairs.

- **Theorem** (Gale-Shapley): Among all stable matchings of n students with n hospitals, this algorithm always finds the unique one that is best possible for every student.

- **Theorem** (Gale-Shapley): Among all stable matchings of n students with n hospitals, this algorithm always finds the unique one that is worst possible for every hospital.

- **Theorem** (Knuth): The Stable Match Algorithm runs in $O(n^2)$ time on input from n students and n hospitals.

Success of the Match Program

NRMP Press Release from March 16, 2018

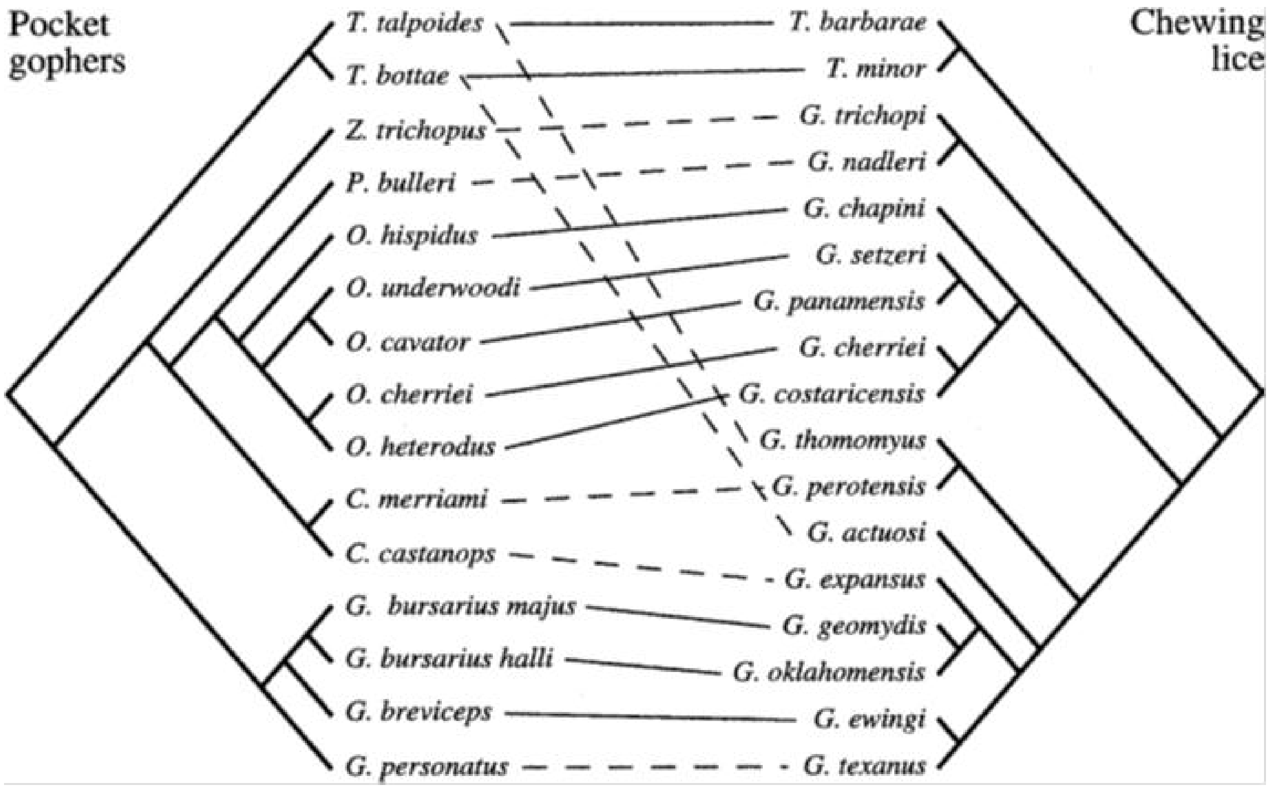
> **<u>Largest Match on Record</u>**
> The 2018 Main Residency Match is the largest in NRMP history. A record-high 37,103 applicants submitted program choices for 33,167 positions, the most ever offered in the Match.

Open Question: What other problems can be solved by the Stable Match Algorithm?

Enumeration Question: How many stable matchings exist for n students and n hospitals? **(https://oeis.org/A005154)**

See Also: The Stable Roommate Problem on Wikipedia

Example 2: Tanglegrams



https://evolutionnews.org/2012/01/parallel_host_a/

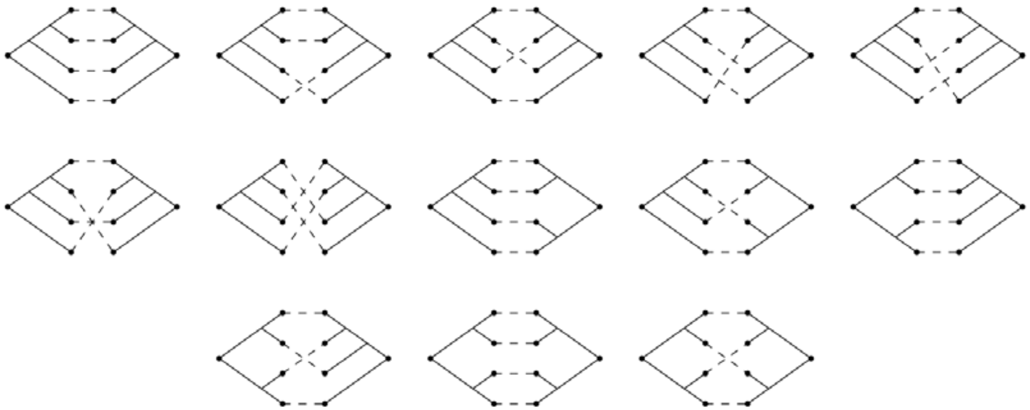
Definition: A tanglegram is a pair of binary trees with a matching between their leaves. They represent two phylogenetic trees of symbiotic organisms.

Counting Tanglegrams

Erick Matsen, Arnold Kas and their team at the Fred Hutchinson Cancer Research Center study mathematical biology.

Enumerative Question (Matsen 2015):
Is there a nice formula to count the number of distinct tanglegrams with n leaves up to symmetries of the left tree and the right tree?

Example: for n=4 there are 13 tanglegrams



Counting Tanglegrams

Enumerative Question (Matsen 2015):
Is there a nice formula to count the number of distinct tanglegrams with n leaves up to symmetries of the left tree and the right tree?

Yes!

Theorem (Billey-Konvalinka-Matsen 2017): The number of tanglegrams of size n is

$$t_n = \sum_{\lambda} \frac{\prod_{i=2}^{\ell(\lambda)} \left(2(\lambda_i + \cdots + \lambda_{\ell(\lambda)}) - 1\right)^2}{z_\lambda},$$

summed over all binary partitions of n. The z-numbers are well known constants.

Counting Tanglegrams

Corollaries of the (Billey-Konvalinka-Matsen) Formula:

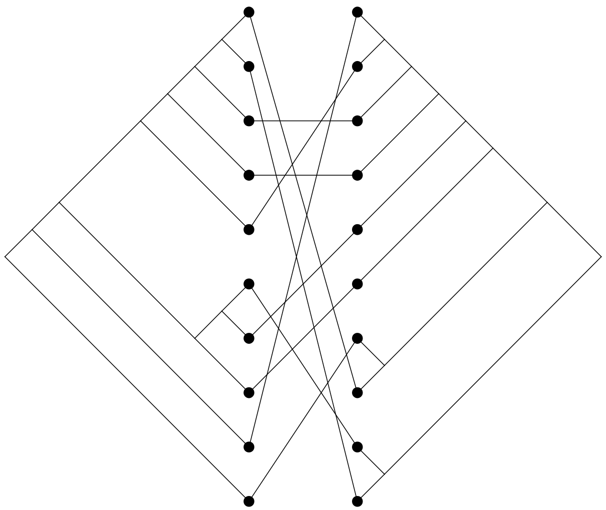
- The number of tanglegrams grows quickly:

$$\frac{t_n}{n!} \sim \frac{e^{\frac{1}{8}} 4^{n-1}}{\pi n^3}$$

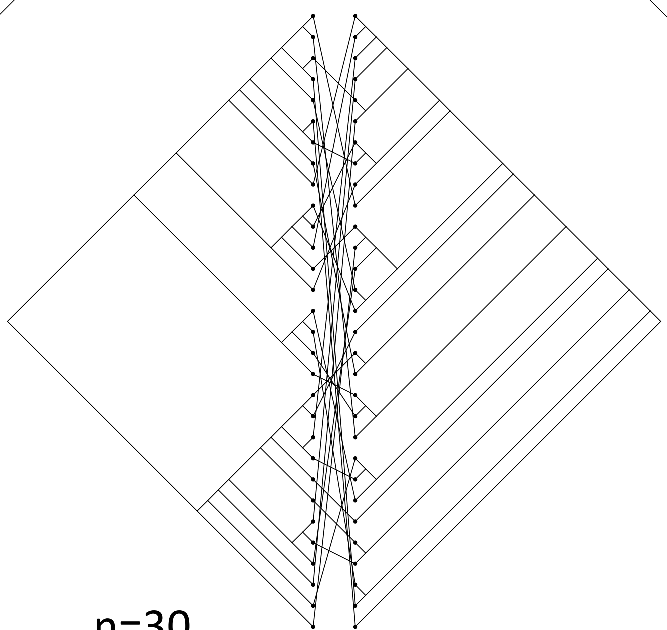
- We can compute the exact number of tanglegrams for n as large as 4000 using a recurrence relation derived from the formula.

- There is an algorithm to find a tanglegram of size n uniformly at random so we can study the average behavior of these objects.

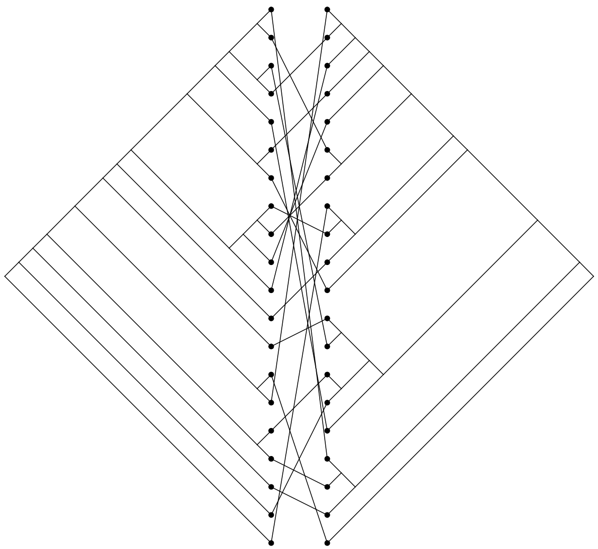
Typical Tanglegrams



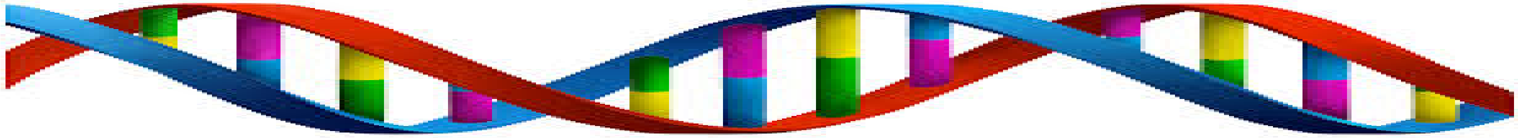
n=10



n=30

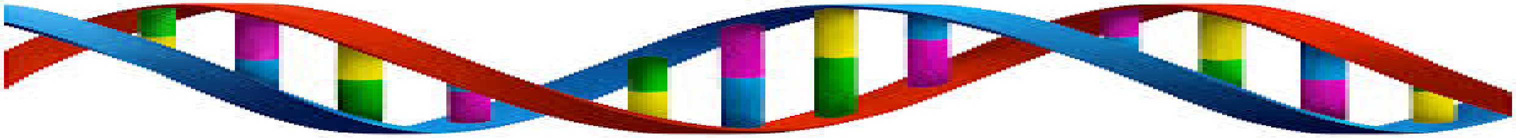


n=20



Reprise

- Questions about existence, enumeration, and optimization of discrete structures appear in many science and industrial applications.
- Combinatorial algorithms for answering these questions have led to faster, cheaper, and more accurate solutions to many problems in our lives.
- Still many questions unanswered.
- Come, join, contribute to the Combinatorics Revolution!



Resources and Acknowledgements

Many thanks to my colleagues on bboard@math.uw for help on preparing this talk!

Thanks to all of you for listening and participating!

Resources:

" College Admissions and the Stability of Marriage". David Gale and Lloyd Shapley. MAA Math Montly 69, 9-14, 1962.

"On the enumeration of tanglegrams and tangled chains" Sara Billey, Matjaz Konvalinka, Frederick A Matsen IV, J. Combin. Theory Ser. A, 146, pp 239--263, 2017.

" Fingerprint Databases for Theorems" Sara Billey and Bridget Tenner. Notices of the AMS 60:8 (2013).

"How to apply de Bruijn graphs to genome assembly" Phillip E C Compeau, Pavel A Pevzner, and Glenn Tesler. Nature Biotechnology 29, 987–991 (2011) http://www.nature.com/nbt/journal/v29/n11/full/nbt.2023.html

MATHEMATICAL FRONTIERS
Combinatorics

Jacques Verstraete,
University of California, San Diego

Graduate Vice-Chair and Professor of Mathematics in the Department of Mathematics at the University of California, San Diego

The Analysis of Finite Structures

Introduction

- Combinatorics is the study of finite structures.
- The central objects in combinatorics are often motivated by concerns in other areas of science, especially theoretical computer science, bioinformatics and statistical physics.
- In turn, combinatorics has applications to other areas of mathematics, such as number theory, geometry, probability, and algebra.

Introduction

- One of the most attractive topics in mathematics is the study of prime numbers.
- Prime factorization: every integer $n > 1$ is a product of primes.

$$11111 = 41 \cdot 271 \qquad 111111 = 3 \cdot 7 \cdot 11 \cdot 37 \qquad 1111111 = 239 \cdot 4649$$

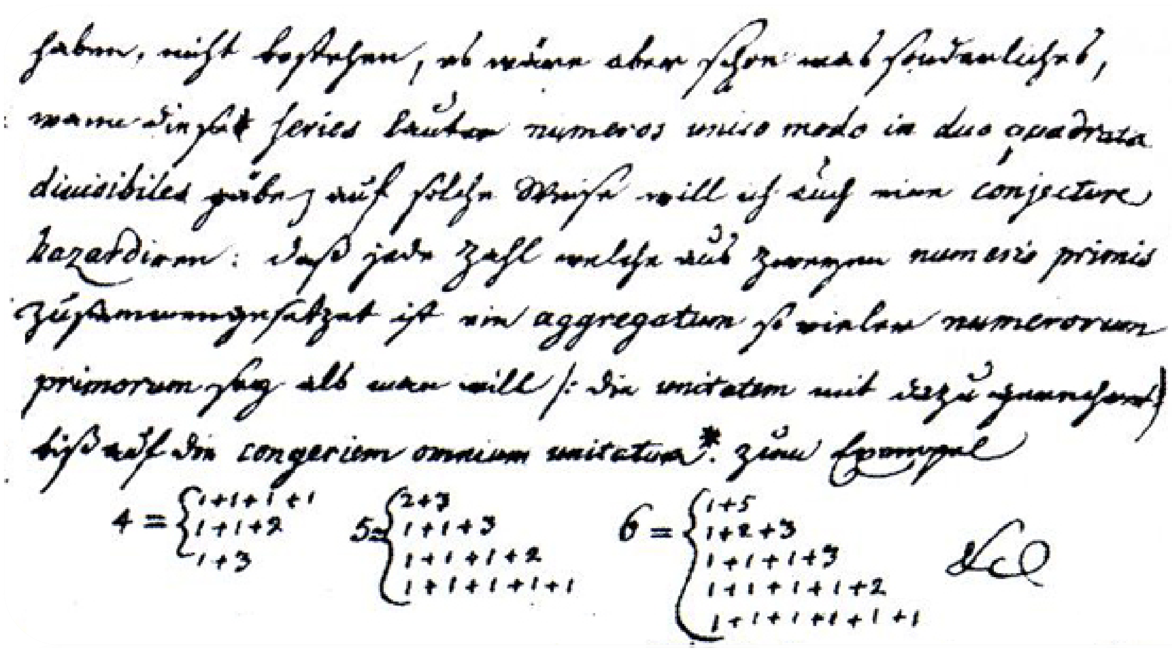
Introduction

- Cryptography: RSA based on hardness of finding prime factorization. Variants underlie much of modern electronic security.

- Primality testing: polynomial time. (Agrawal, Kayal, Saxena, 2002)

Introduction

- Goldbach's Conjecture[1742] : every even number larger than 2 is the sum of two primes.


divisibiles giebt, auf solche Weise will ich auch eine conjecture
hazardiren: dass jede Zahl welche aus zweyen numeris primis
zusammengesetzt ist ein aggregatum so vieler numerorum
primorum sey als man will (die unitatem mit dazu gerechnet)
bis auf die congeriem omnium unitatum.

$4 = \begin{cases} 1+1+1+1 \\ 1+1+2 \\ 1+3 \end{cases}$ $5 = \begin{cases} 2+3 \\ 1+1+3 \\ 1+1+1+2 \\ 1+1+1+1+1 \end{cases}$ $6 = \begin{cases} 1+5 \\ 1+2+3 \\ 1+1+1+3 \\ 1+1+1+1+2 \\ 1+1+1+1+1+1 \end{cases}$

Quasirandomness

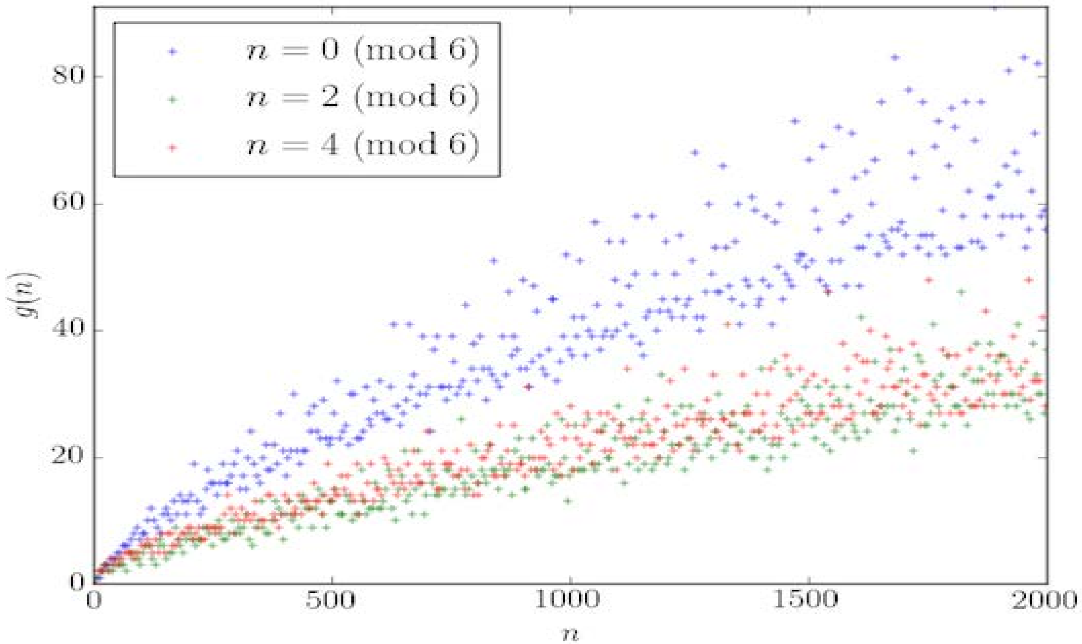
- The Extended Goldbach Conjecture states that the number $R(n)$ of representations of n as a sum of two primes satisfies:

$$R(n) \sim 2\Pi_2 \quad \cdot \quad \prod_{\substack{p>2 \\ p|n}} \frac{p-1}{p-2} \quad \cdot \quad \int_2^n \frac{1}{(\log x)^2}\, dx$$

(Hardy-Littlewood, 1923)

Introduction

- Goldbach's Conjecture[1742] : every even number larger than 2 is the sum of two primes.

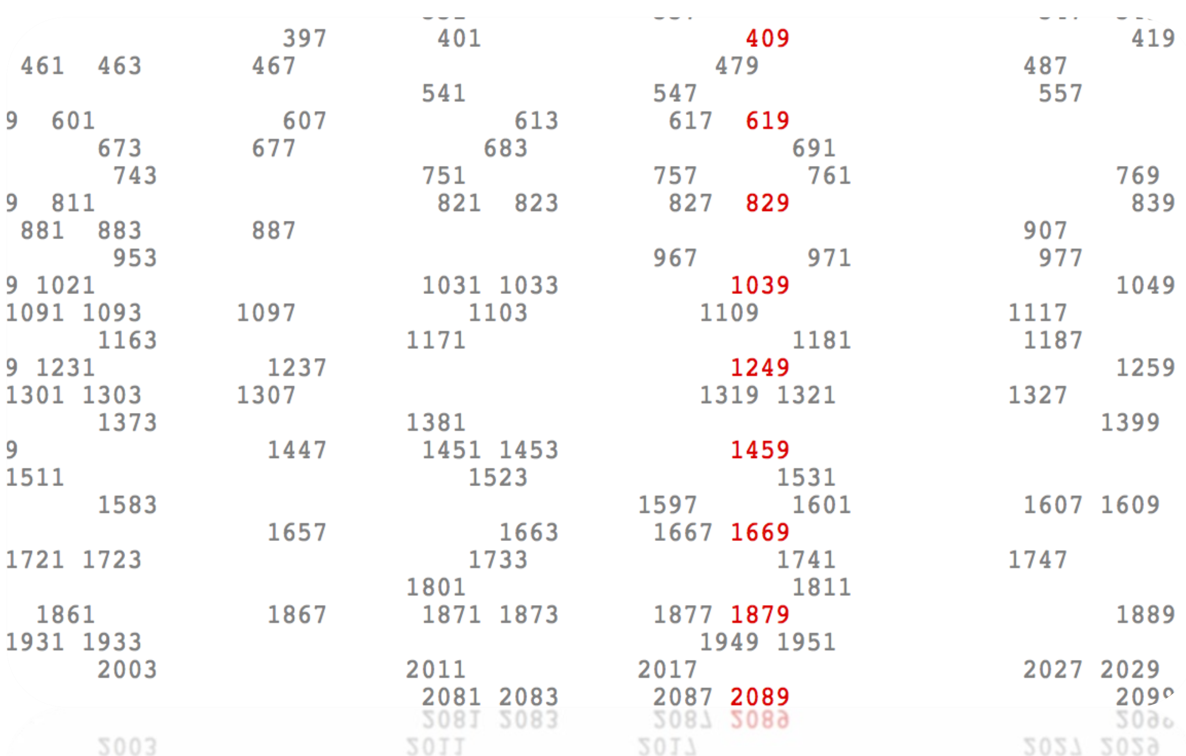


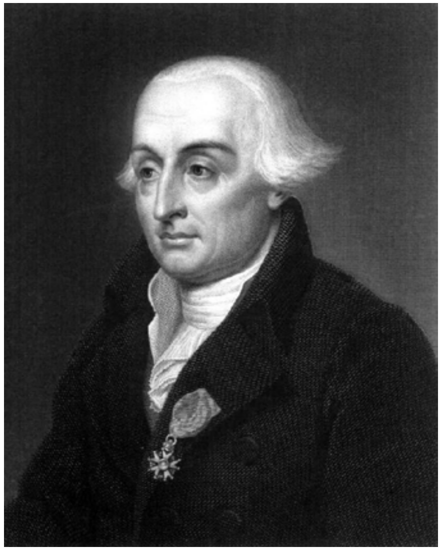
$$R(22) = 3$$

$$\begin{aligned} 22 &= 3 + 19 \\ &= 5 + 17 \\ &= 11 + 11 \end{aligned}$$

Introduction

- Lagrange[1770] : for every positive integer k there exists a progression of k primes.





The probabilistic method

- In many areas of mathematics, one is required to construct a structure under a prescribed list of constraints, or at least prove its existence.

- The probabilistic method was introduced by Paul Erdõs over fifty years ago.

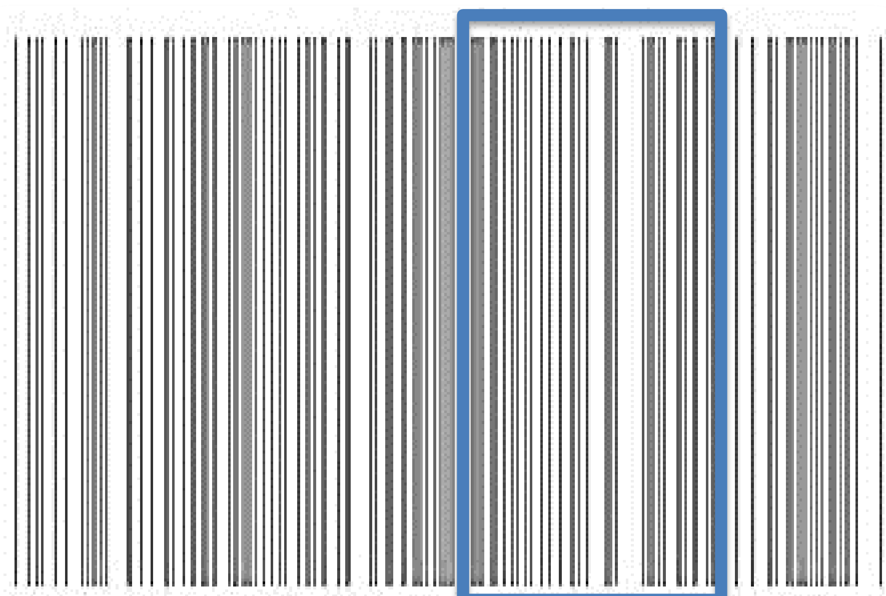
- The next examples illustrate one of the organizing principles of the method:

if it seems likely that the structure we want is roughly uniform,

then a random example is worth trying.

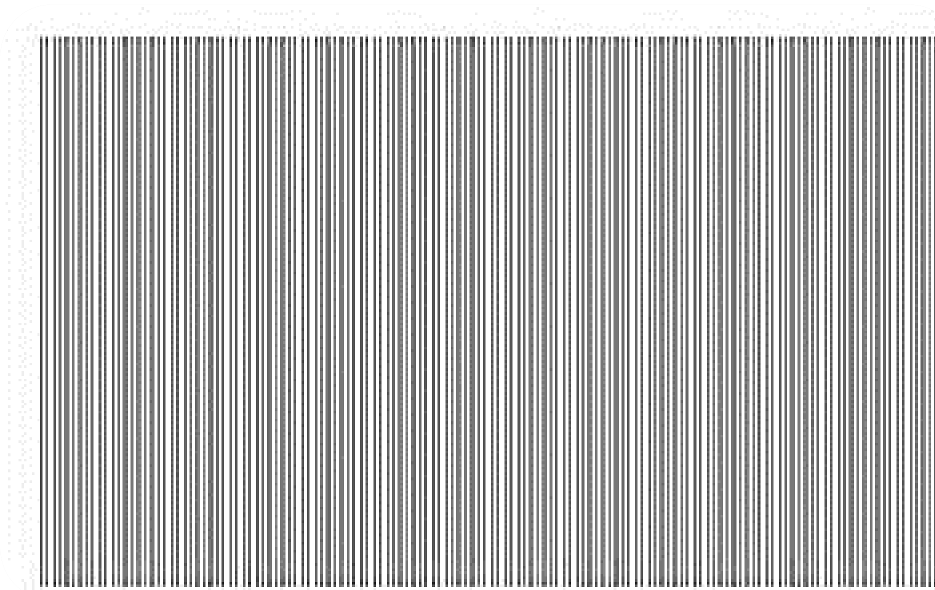
Randomness versus Structure

- Suppose we select a random set of numbers from 1 to n, where each number is selected independently with probability p.
- We would expect every interval of m consecutive numbers contains about pm selections.



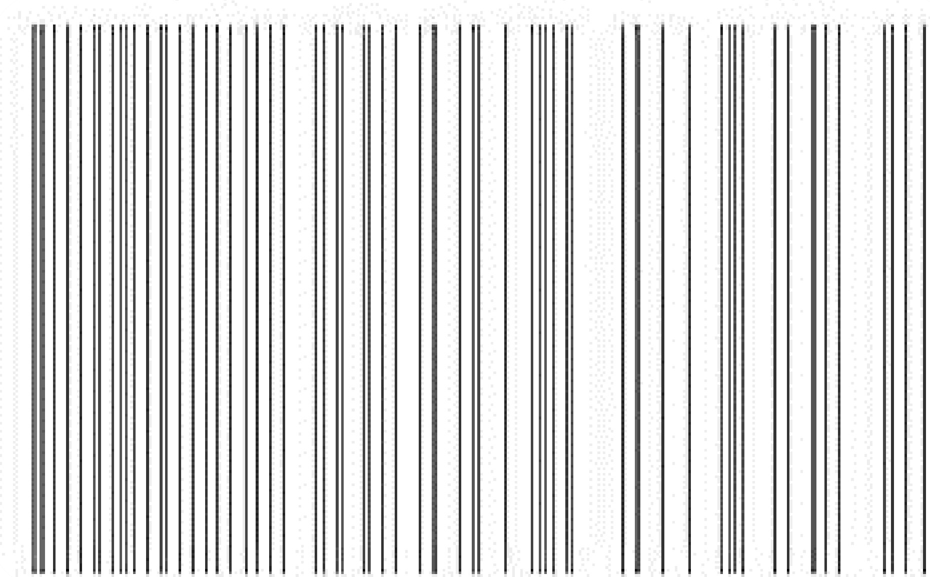
Randomness versus Structure

- The set of even numbers, on the other hand, should be considered to be “structured”.
- More generally, any union of few arithmetic progressions should be considered “structured”



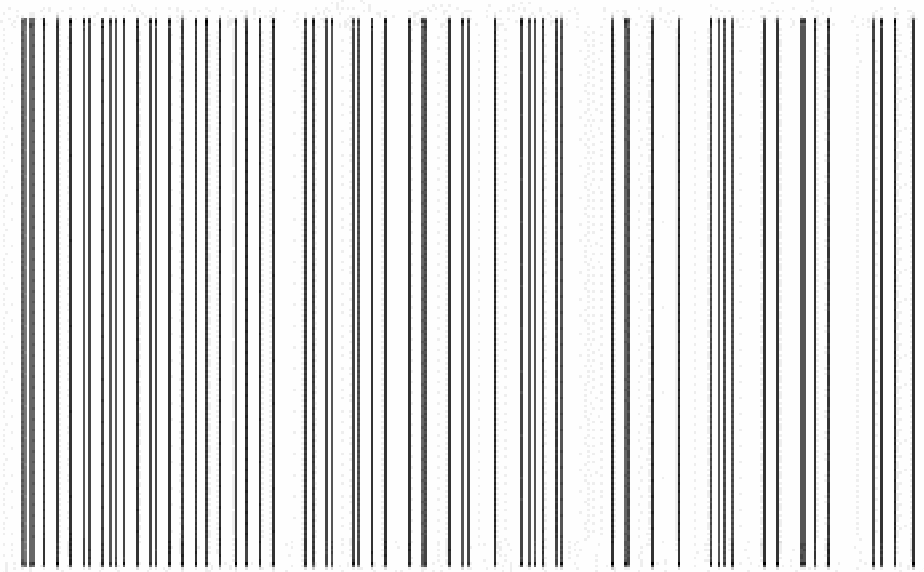
Randomness versus Structure

- For instance, consider the set of prime numbers.
- According to the Prime Number Theorem, there are roughly $\frac{n}{\log n}$ primes less than n.



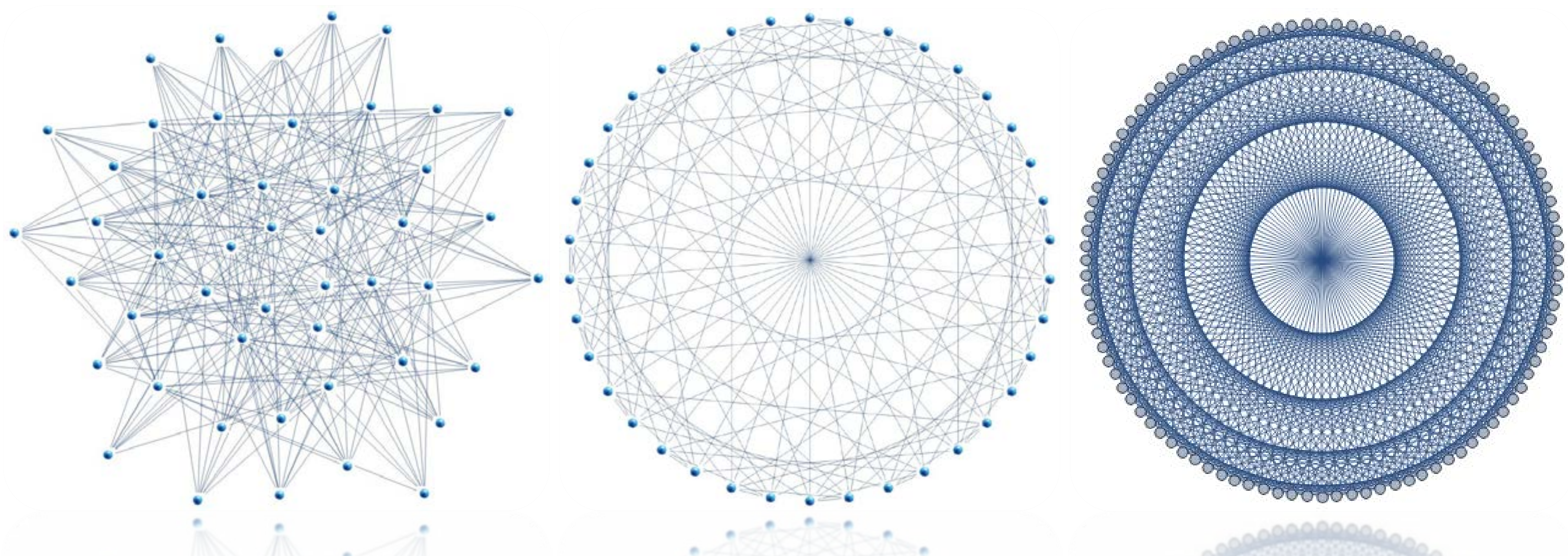
Randomness versus Structure

- Cramér's Conjecture : There is a prime between n and about $n + (\log n)^2$ for every n. (Cramér, 1936)



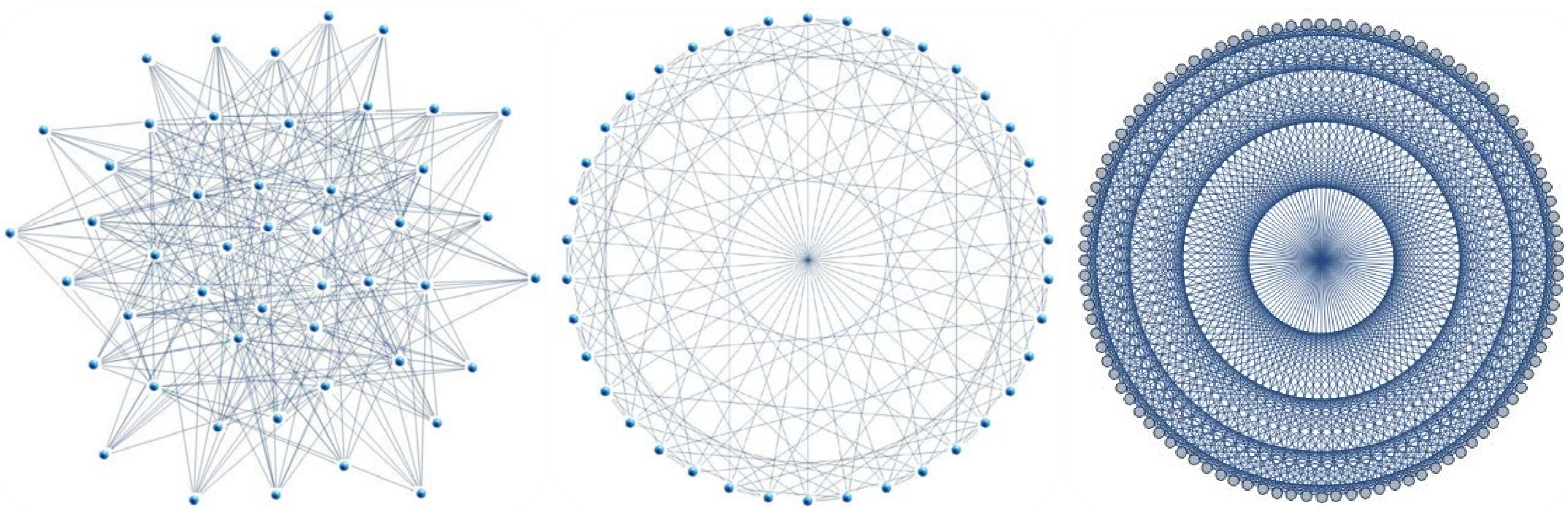
Randomness versus Structure

- A graph is a set of vertices / nodes together with a set of pairs of vertices called edges.
- These are fundamental objects in combinatorics.



Randomness versus Structure

- When is a graph "random"?
- Place edges randomly and independently with probability p.



Quasirandomness

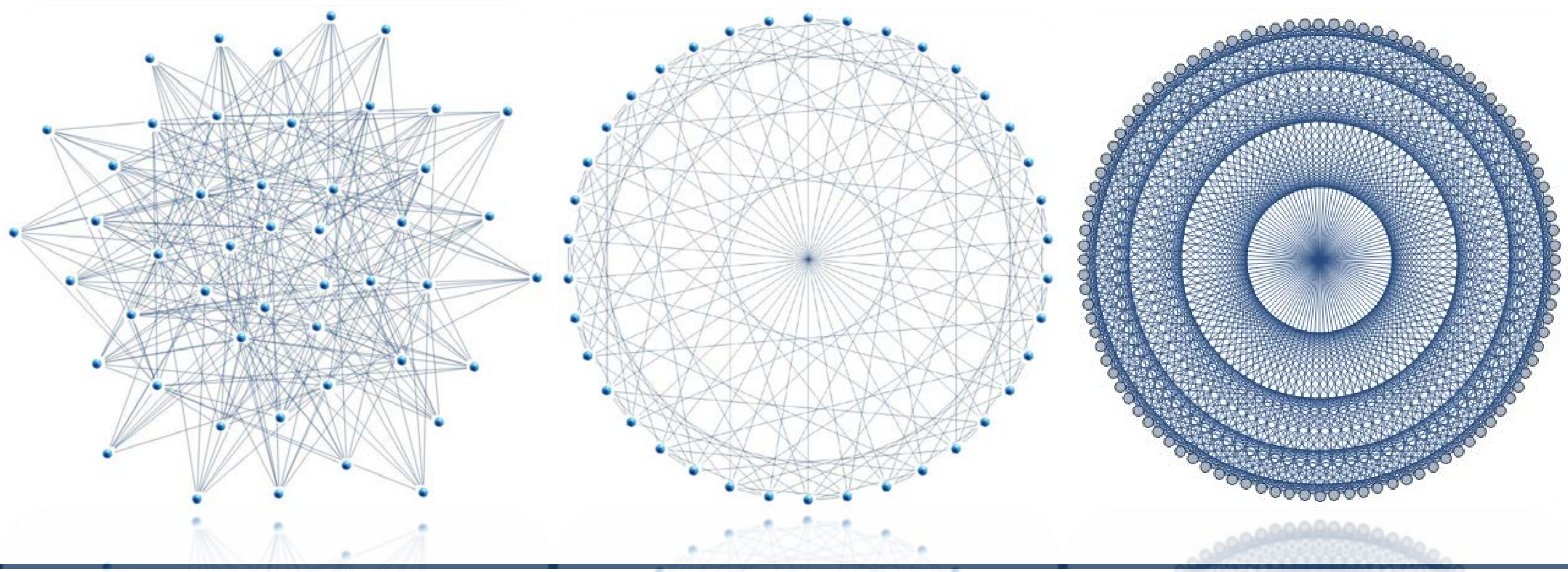
- Given any set X of vertices, we expect $p\binom{|X|}{2}$ edges of the graph to lie inside X.

- We call an n-vertex graph of density p an ε-quasirandom graph if for every set X

$$\left| e(X) - p\binom{|X|}{2} \right| < \varepsilon p n^2$$

Randomness versus Structure

- When is a graph "quasirandom"?



Quasirandomness

- How to tell if a graph is random? Using spectral theory of the graph matrices.

- Expander Mixing Lemma (Alon, 1986)

$$\left| e(X) - p\binom{|X|}{2} \right| \le \lambda |X|$$

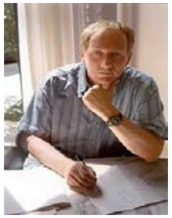
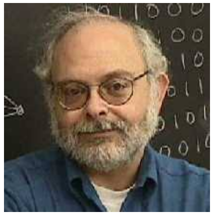
Quasirandomness

- How to tell if a graph is random? Counting quadrilaterals.
- Thomason (1987), Chung-Graham-Wilson (1991)

A graph with n vertices and density p is ε-quasirandom if and only if the number of quadrilaterals in the graph is at most $(1+\varepsilon^4)(pn)^4$.

- Quasirandom graphs appear frequently in applications, for example in coding and information theory (expander graphs).



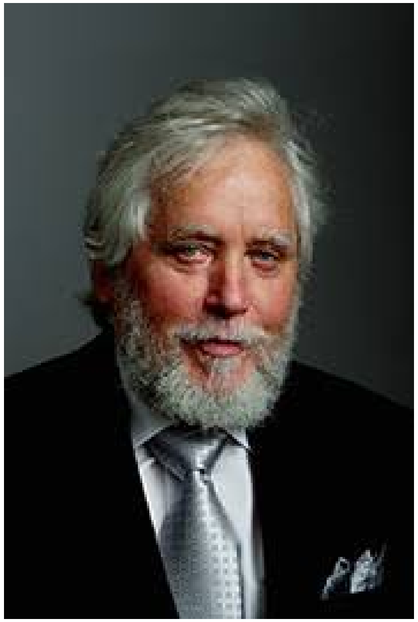



Quasirandomness

- We can use graphs to find arithmetic progressions in sets of integers.

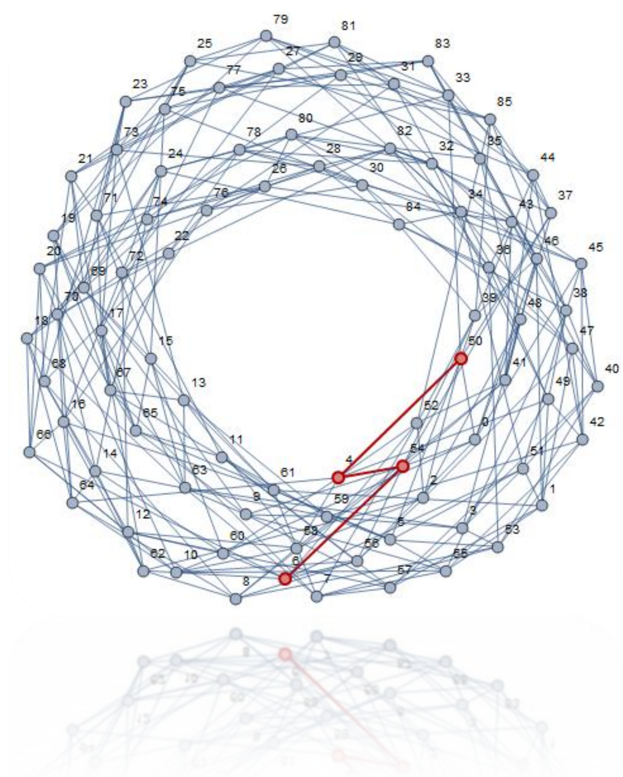
- Szemeredi's Theorem (1975)

  Every set of integers positive density contains arbitrarily long progressions.



Quasirandomness

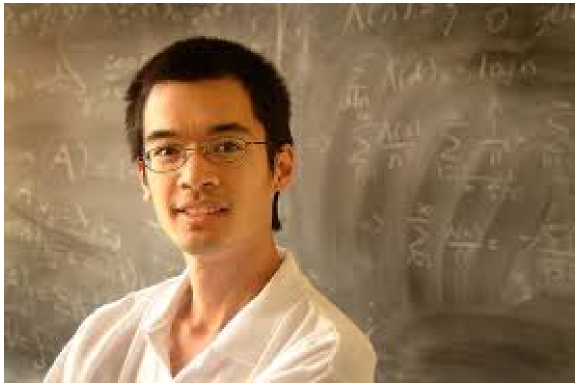
- The arithmetic progression {3,5,7}



Breakthroughs

- Theorem. (Green-Tao Theorem, 2006)

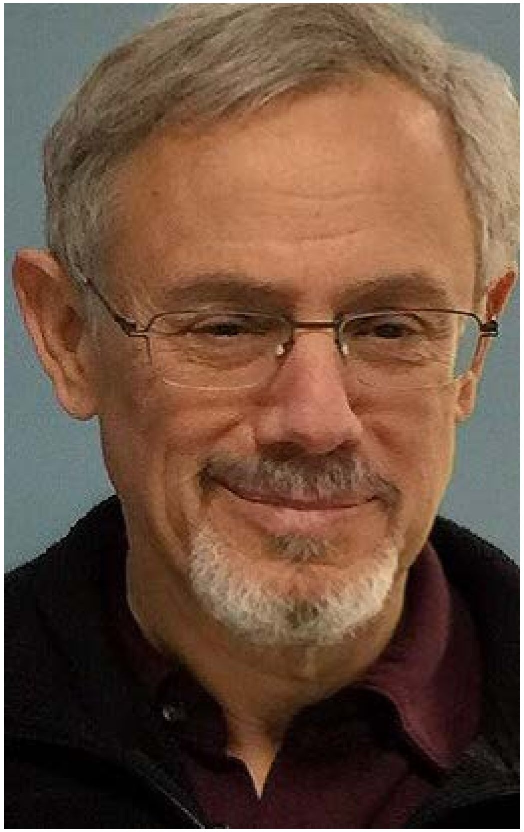
  The primes contain arbitrarily long arithmetic progressions.



Conclusion

- Combinatorics has burgeoned into a fundamental part of modern mathematics, establishing many connections and applications to many other areas of science.
- We discussed a general modern theme in combinatorics, which is to distinguish between randomness and structure in combinatorial objects.
- The probabilistic method has led to a number of recent breakthroughs.

MATHEMATICAL FRONTIERS
Combinatorics



Sara Billey,
University of Washington

Jacques Verstraete,
University of California, San Diego

Mark Green,
UCLA (moderator)

View webinar videos and learn more about BMSA at www.nas.edu/MathFrontiers

MATHEMATICAL FRONTIERS

2018 Monthly Webinar Series, 2-3pm ET

February 13*:
Mathematics of the Electric Grid

March 13*:
Probability for People and Places

April 10*:
Social and Biological Networks

May 8*:
Mathematics of Redistricting

June 12*: *Number Theory: The Riemann Hypothesis*

July 10*: *Topology*

August 14*: *Algorithms for Threat Detection*

September 11*: *Mathematical Analysis*

October 9: *Combinatorics*

November 13:
Why Machine Learning Works

December 11:
Mathematics of Epidemics

***** *Recording posted***

Made possible by support for BMSA from the National Science Foundation Division of Mathematical Sciences and the Department of Energy Advanced Scientific Computing Research

View webinar videos and learn more about BMSA at www.nas.edu/MathFrontiers