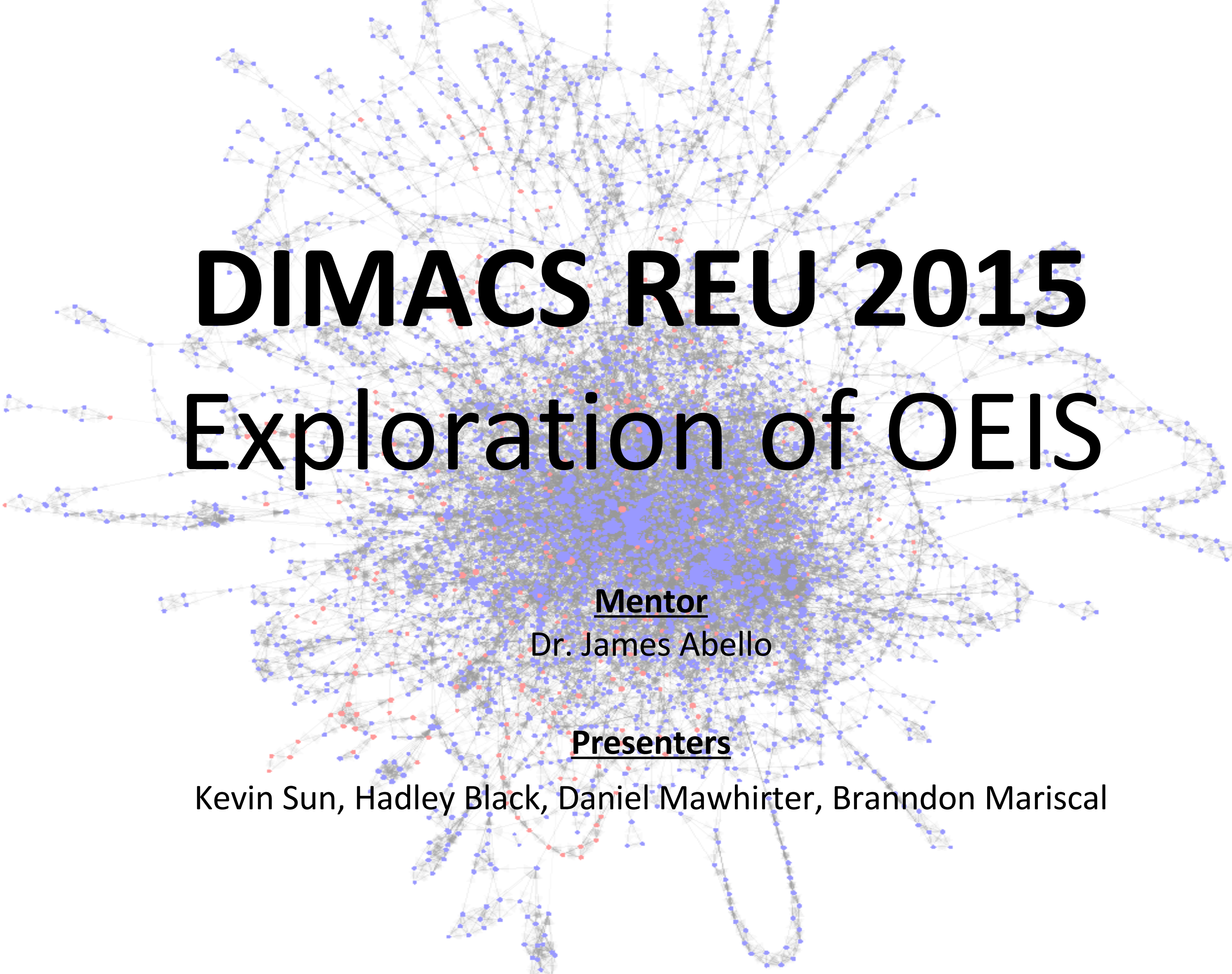


DIMACS REU 2015

Exploration of OEIS

Mentor

Dr. James Abello

Presenters

Kevin Sun, Hadley Black, Daniel Mawhirter, Branndon Mariscal

Project Goals

- To develop an application that enhances users’ experience with OEIS
- To create a k-partite multigraph representation of the OEIS database, to facilitate the process of filtering data
- To promote interest in the OEIS and the general applicability of the Graph Card abstraction

Outline

- Goals of the project
- Description of the project
- Stumbling blocks
- Sample findings
  - Clustering
  - Implementation
- Summary
- Ideas for extensions

Description

- Develop graphical representation of OEIS: The On-Line Encyclopedia of Integer Sequences
- Explore various relationships among sequences
- Expand user capabilities when navigating OEIS

Flow of Assignments

- Hadley - provide clustering of large graphs
- Branndon - analyze graphs and data
- Daniel - develop user interface, maintain database
- Kevin - label graphs, implement non-clustering ideas

THE ON-LINE ENCYCLOPEDIA OF INTEGER SEQUENCES®

founded in 1964 by N. J. A. Sloane

Search Hints

(Greetings from The On-Line Encyclopedia of Integer Sequences!)

A233743 7*binomial(6*n+7, n)/(6*n+7).

1, 7, 63, 644, 7105, 82467, 992446, 12271512, 154962990, 1990038435, 25909892008, 341225775072, 4537563627415, 60842326873230, 821692714673340, 11167153485624304, 152610018401940330, 2095863415900961490, 28910564819681953485, 400379714692751795820 (list; graph; refs; listen; history; text; internal format)

OFFSET	0,2
COMMENTS	Fuss-Catalan sequence is a(n,p,r) = r*binomial(np+r,n)/(np+r), this is the case p=6, r=7.
LINKS	Vincenzo Librandi, Table of n, a(n) for n = 0..200 J-C. Aval, Multivariate Fuss-Catalan Numbers, arXiv:0711.0906v1, Discrete Math., 308 (2008), 4660-4669. Thomas A. Dowling, Catalan Numbers Chapter 7 Wojciech Mlotkowski, Fuss-Catalan Numbers in Noncommutative Probability, Docum. Mathm. 15: 939-955.
FORMULA	G.f. satisfies: B(x) = {1 + x*B(x)^(p/r)}^r, where p=6, r=7.
MATHEMATICA	Table[7 Binomial[6 n + 7, n]/(6 n + 7), {n, 0, 40}] (* Vincenzo Librandi, Dec 16 2013 *)
PROG	(PARI) a(n) = 7*binomial(6*n+7, n)/(6*n+7); (PARI) {a(n)=local(B=1); for(i=0, n, B=(1+x*B^(6/7))^7+x*O(x^n)); polcoeff(B, n)} (MAGMA) [7*Binomial(6*n+7, n)/(6*n+7): n in [0..30]]; // Vincenzo Librandi, Dec 16 2013
CROSSREFS	Cf. A000108, A002295, A212071, A212072, A212073, A130564, A233827, A233829, A233830. Sequence in context: A218237 A024088 A155132 * A015684 A051579 A185106 Adjacent sequences: A233740 A233741 A233742 * A233744 A233745 A233746
KEYWORD	nonn
AUTHOR	Tim Fulford, Dec 15 2013
EXTENSIONS	More terms from Vincenzo Librandi, Dec 16 2013
STATUS	approved

Initial Findings (1)

A242533: *Number of cyclic arrangements of S={1,2,...,2n} such that the difference of any two neighbors is coprime to their sum.*

A242530: *Number of cyclic arrangements of S={1,2,...,2n} such that the binary expansions of any two neighbors differ by one bit.*

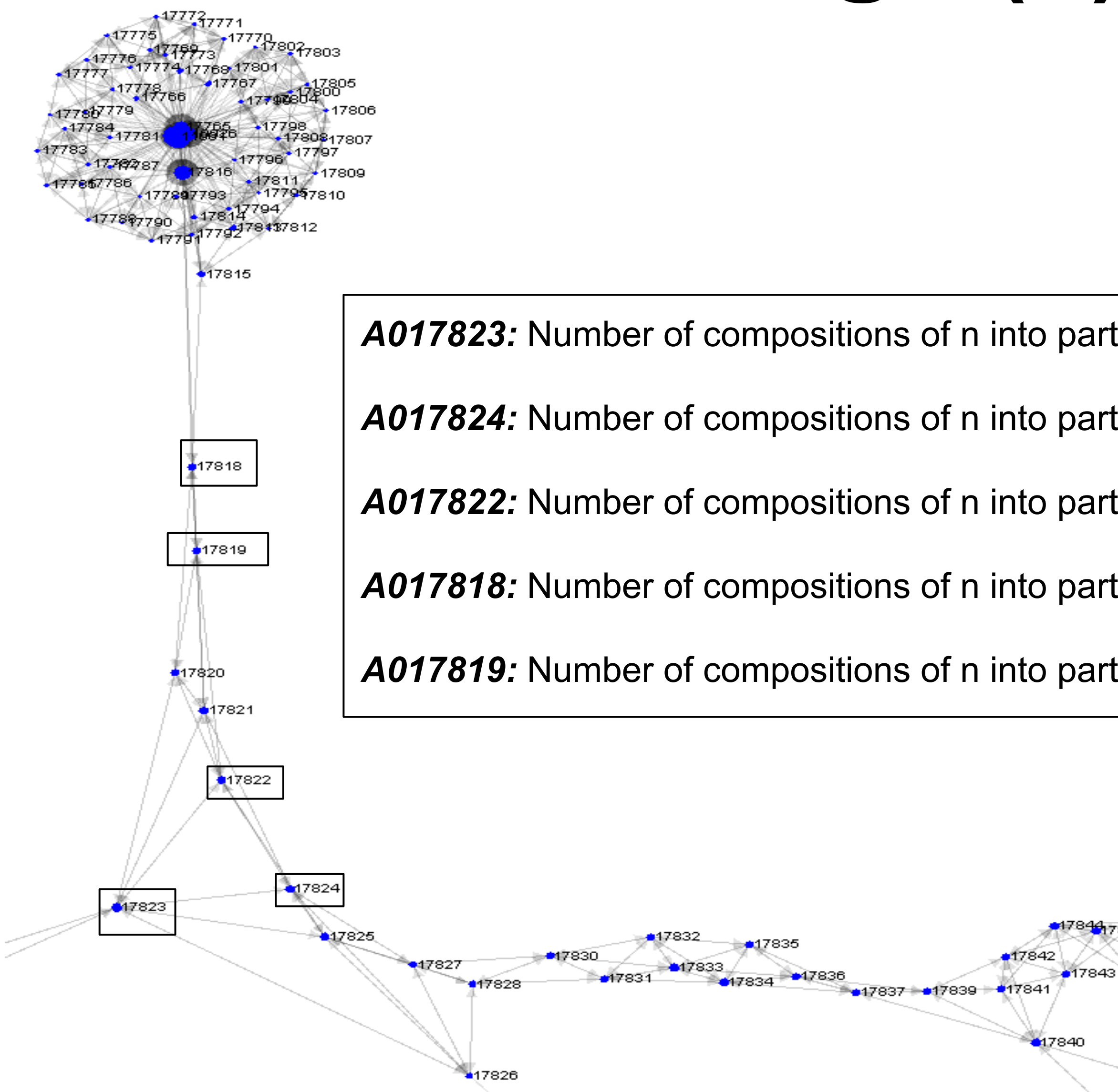
A242521: *Number of cyclic arrangements (up to direction) of {1,2,...,n} such that the difference between any two neighbors is b^k for some b>1 and k>1.*

A074426: *Number of 6-ary Lyndon words of length n with trace 0 and subtrace 4 over Z_6.*

A074438: *Number of 6-ary Lyndon words of length n with trace 2 and subtrace 4 over Z_6.*

A074424: *Number of 6-ary Lyndon words of length n with trace 0 and subtrace 2 over Z_6.*

Initial Findings (2)



A017823: Number of compositions of n into parts p where 3 <= p <= 10.

A017824: Number of compositions of n into parts p where 3 <= p <= 11.

A017822: Number of compositions of n into parts p where 3 <= p <= 9.

A017818: Number of compositions of n into parts p where 3 <= p <= 5.

A017819: Number of compositions of n into parts p where 3 <= p <= 6.

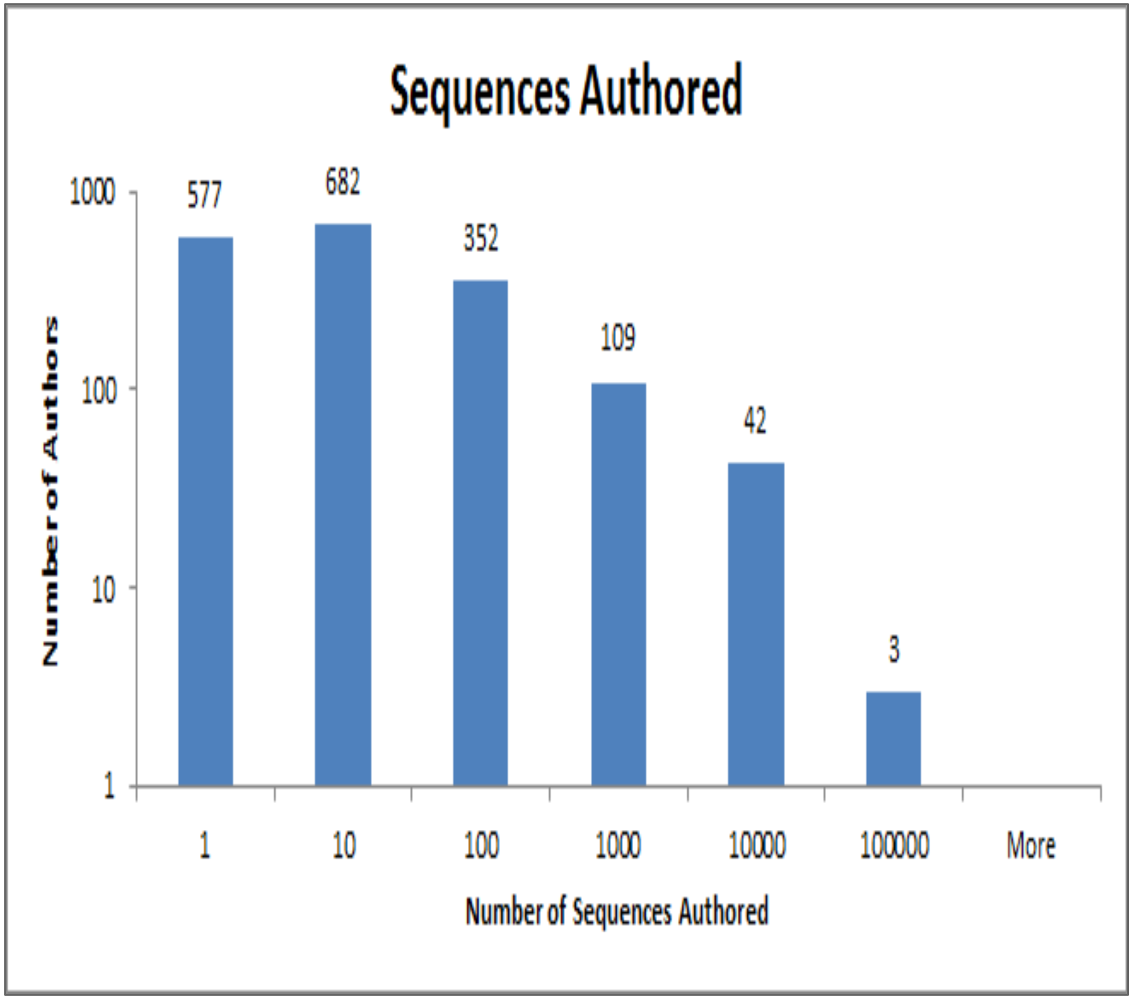
Statistical Findings - Authors

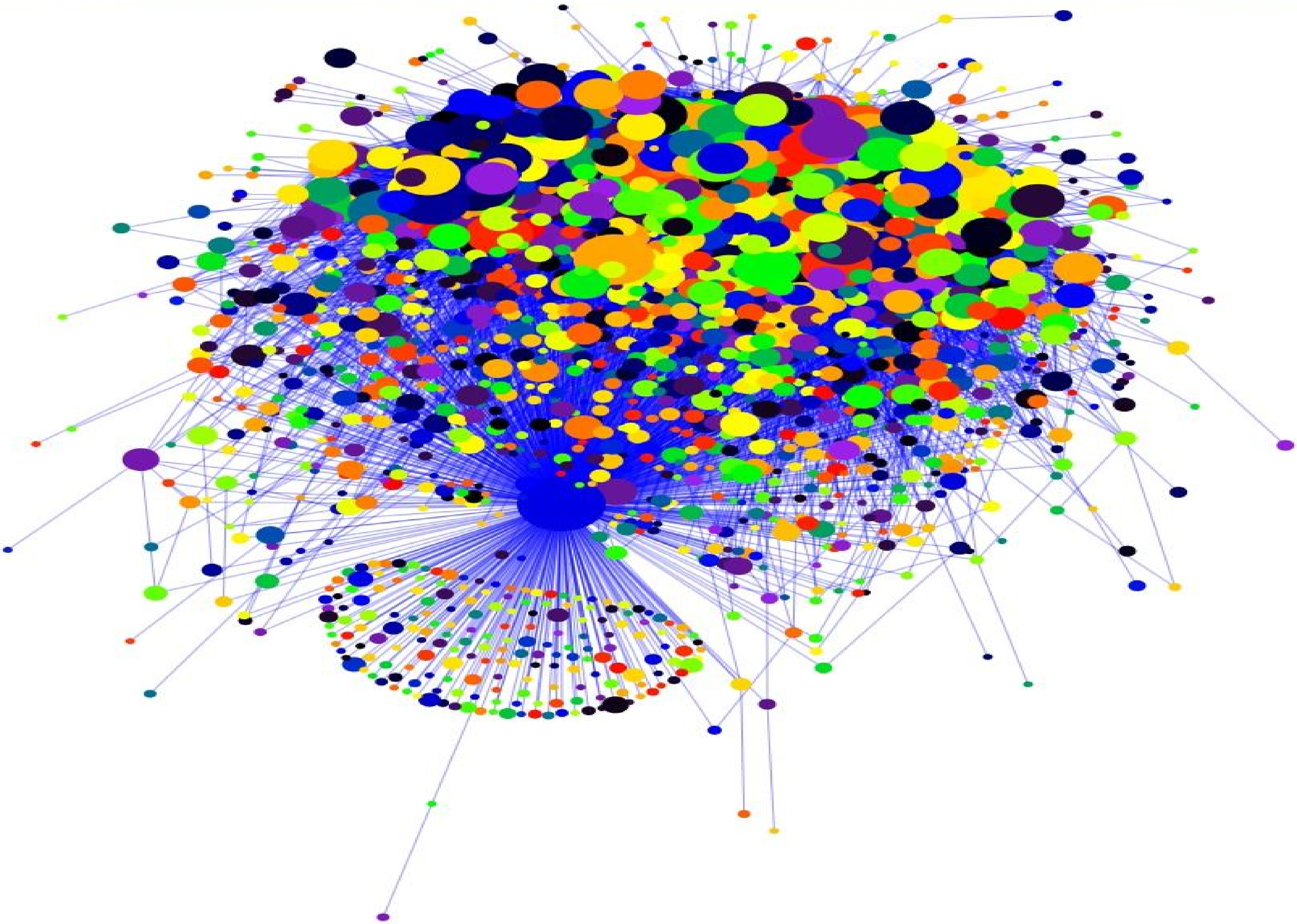
Statistics

- Mean: 136.6074
- Standard Deviation: 1,330.433
- Median: 3
- 1,787 authors span 96.5% of OEIS (249,521 sequences)

Top 5 Authors

Name	Description	Number of Sequences Authored
N.J.A. Sloane	Creator of OEIS	39,867
R.H. Hardin	Programmer	30,720
C. Kimberling	University of Evansville, Indiana. (Geometry)	17,442
P.D. Hanna		6,544
Amarnath Murthy		5,274

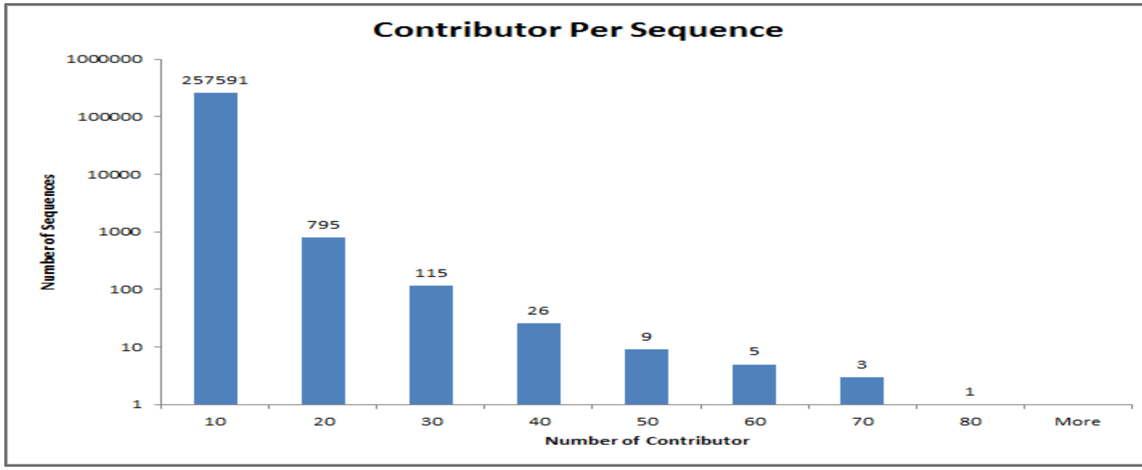




Statistical Findings - Contributors per Sequence

Statistics

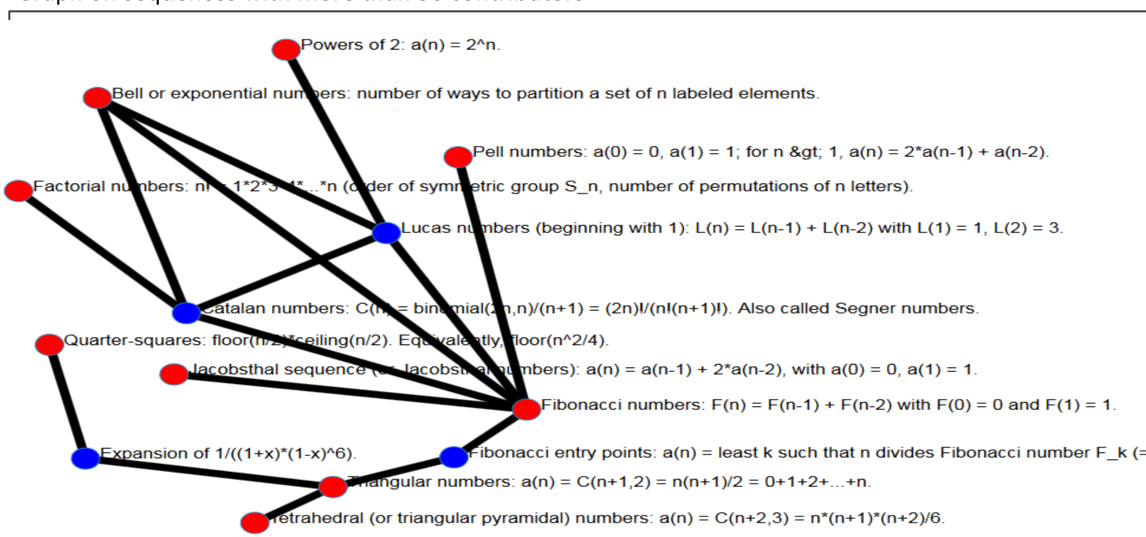
- Mean: 1.741921
- Standard Deviation: 1.570046
- Median: 1
- Min: 0
- Max: 72



Top 5 Sequences

Sequence ID	Title	Number of Contributors
A45	Fibonacci numbers	72
A217	Triangular numbers	67
A292	Tetrahedral numbers	61
A79	Powers of 2	61
A129	Pell numbers	59

Graph on sequences with more than 50 contributors



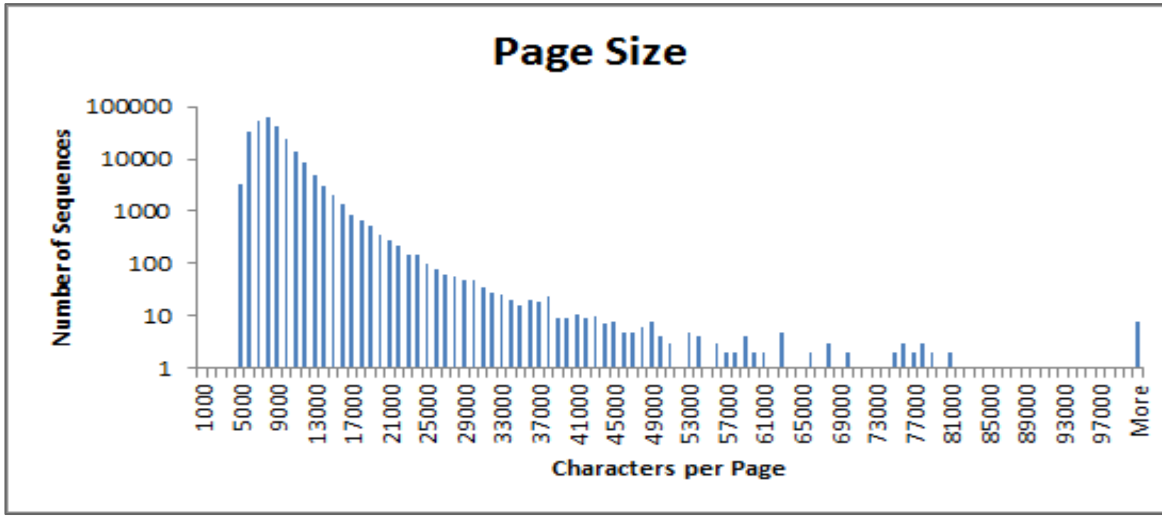
Statistical Findings - Page Size

Statistics

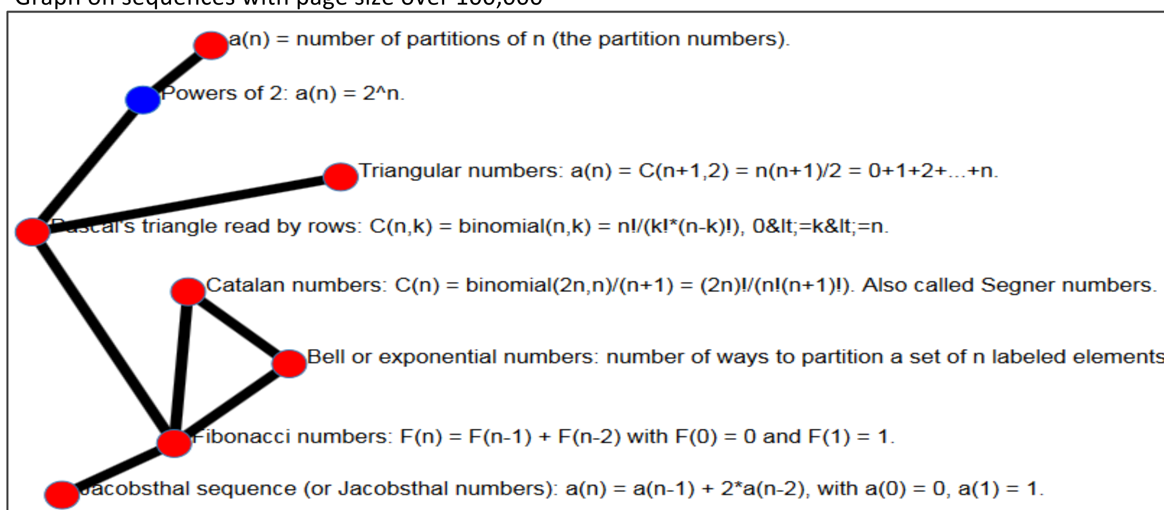
- Mean: 8,091
- Standard Deviation: 2,827
- Median: 7,565
- Min: 4,215
- Max: 170,085

Top 5 Sequences

Sequence ID	Title	Characters per Page
A108	Catalan numbers	170,085
A45	Fibonacci numbers	168,644
A110	Bell or exponential numbers	142,159
A217	Triangular numbers	108,838
A41	a(n) = number of partitions of n	104,831



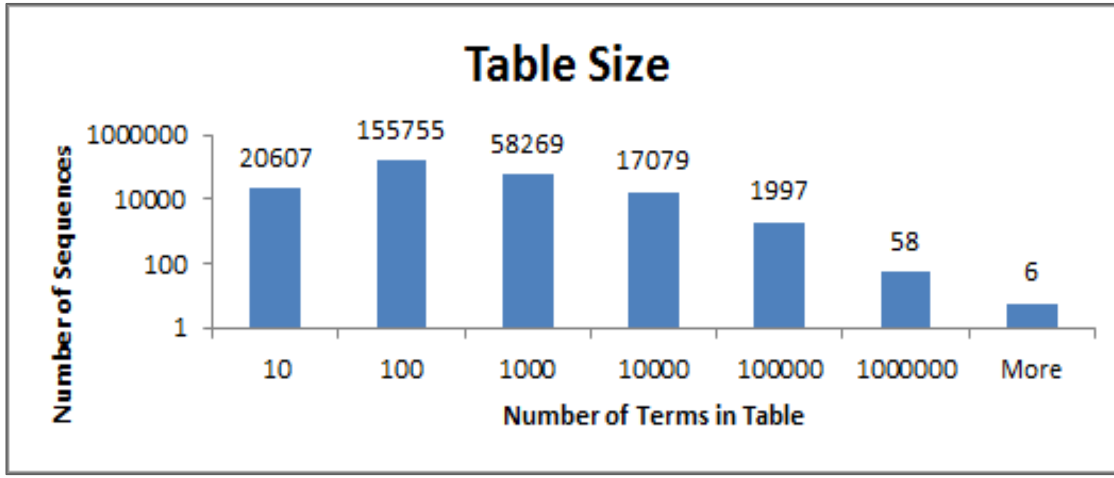
Graph on sequences with page size over 100,000



Statistical Findings - Table Size

Statistics

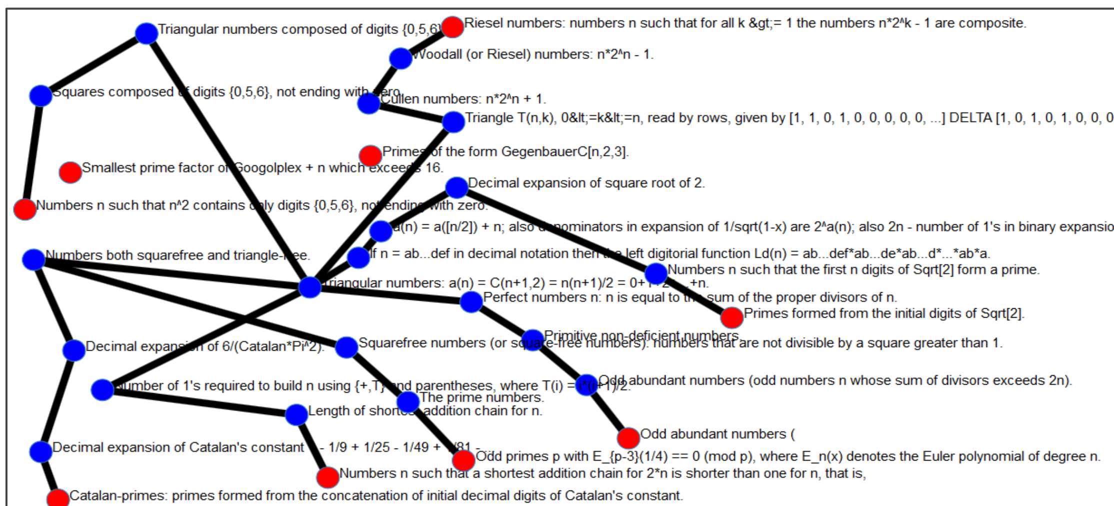
- Mean: 1,008.6037
- Standard Deviation: 12,660.09
- Median: 53
- Min: 1
- 8% of OEIS sequences have 10 or less terms given (20,607 sequences)



Exemplary Sequences

Sequence ID	Title	Number of terms in table
A58471	Numbers n such that n^2 contains only digits {7,8,9}	2
A217228	Number of triangular nXnXn arrays of occupancy after each element moves to some neighbor but without 2-loops and with no occupancy greater than 2.	6

Graph on sequences with only one term in table



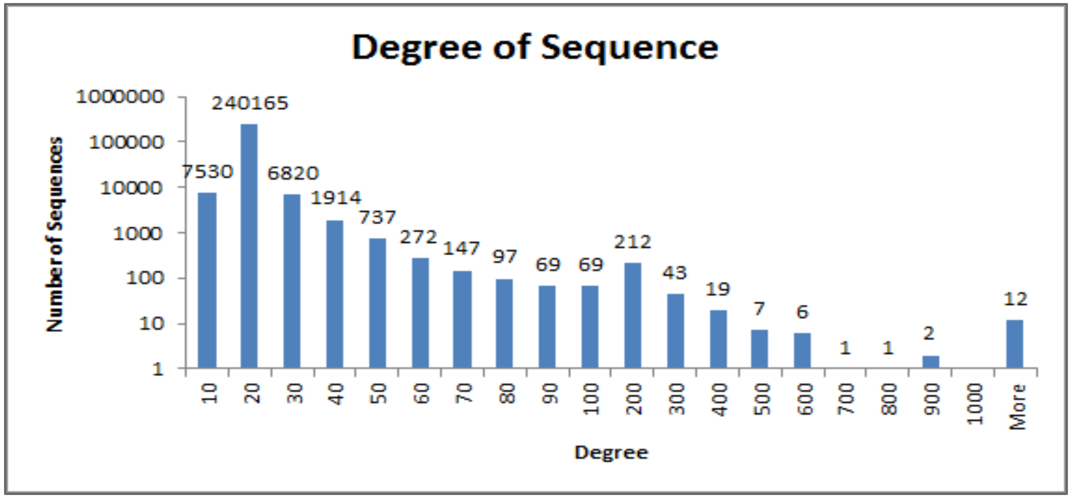
Statistical Findings - Degree

Statistics

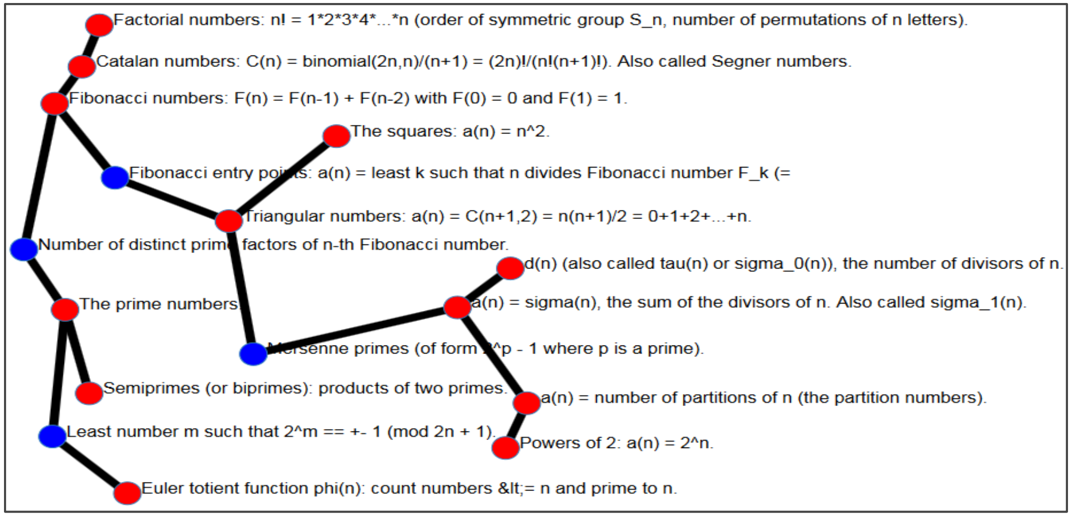
- Mean: 21.78595
- Standard Deviation: 17.7459
- Median: 20
- 93% of OEIS sequences have degree 10 through 20. (240,165 sequences)

Top 5 Sequences

Sequence ID	**Title**	**Degree**
A40	The primes	5,482
A45	Fibonacci numbers	2,367
A217	Triangular numbers	1,903
A203	Sigma(n) = The sum of divisors of n	1,733
A10	Euler totient function phi(n): count numbers <= n and prime to n	1,507



Graph on sequences with degree over 1,000



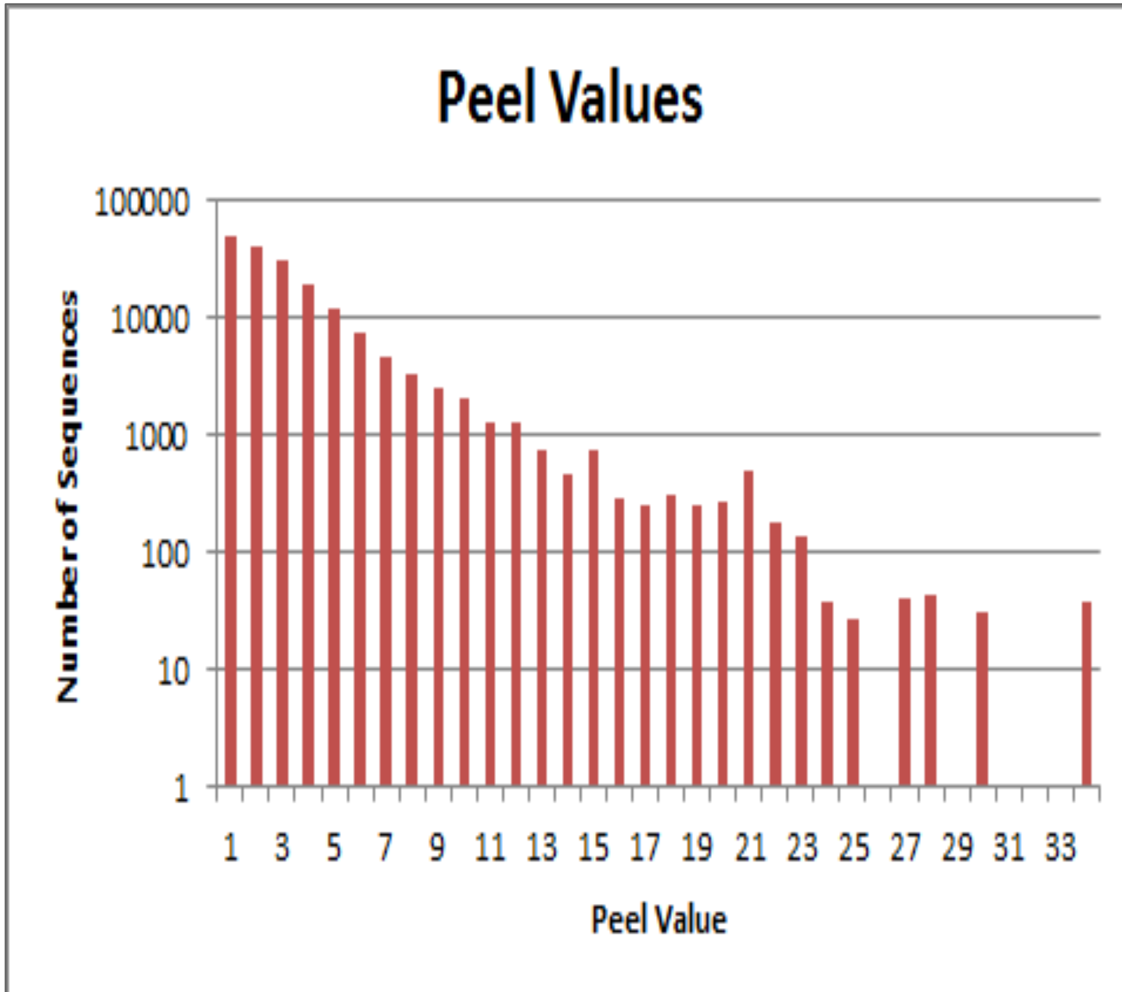
Statistical Findings - Peel Value

Statistics

- Mean: 5046.5
- Standard Deviation: 11239.51
- Median: 288

Exemplary Sequences

Sequence ID	**Title**	**Peel Value**
A78457	a(n) = least positive k such that the remainder when 3^k is divided by k is n.	34
A128149	Least k such that n^k mod k = n-1.	34
A257377	Numbers n such that n, n+2, n+6, n+12, n+14, n+20, n+24, n+26, n+30, n+36, n+42, n+44, n+50, n+54, n+56, n+62 and n+66 are all prime.	30
A27569	Initial members of prime decaplets (p, p+2, p+6, p+8, p+12, p+18, p+20, p+26, p+30, p+32)	28
A34851	Rows of Losanitsch's triangle T(n, k), n >= 0, 0 <= k <= n.	28



Peel 34 Graph

A128158: a(n) = least k such that the remainder when 18^k is divided by k is n.

A128159: a(n) = least k such that the remainder when 19^k is divided by k is n.

A128159: a(n) = least k such that the remainder when 28^k is divided by k is n.

A127817: a(n) = least k such that the remainder when 9^k is divided by k is n.

A128157: a(n) = least k such that the remainder when 17^k is divided by k is n.

A119715: a(n) = least k such that the remainder when 7^k is divided by k is n.

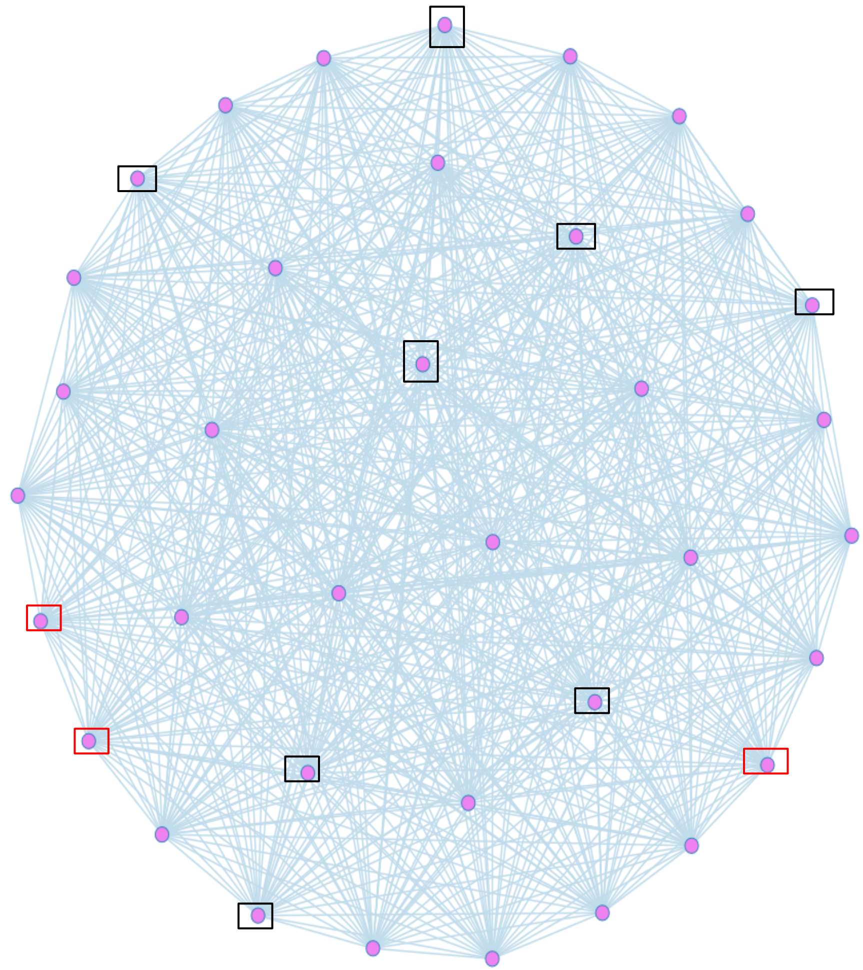
A127820: a(n) = least k such that the remainder when 12^k is divided by k is n.

A119678: a(n) = least k such that the remainder when 4^k is divided by k is n.

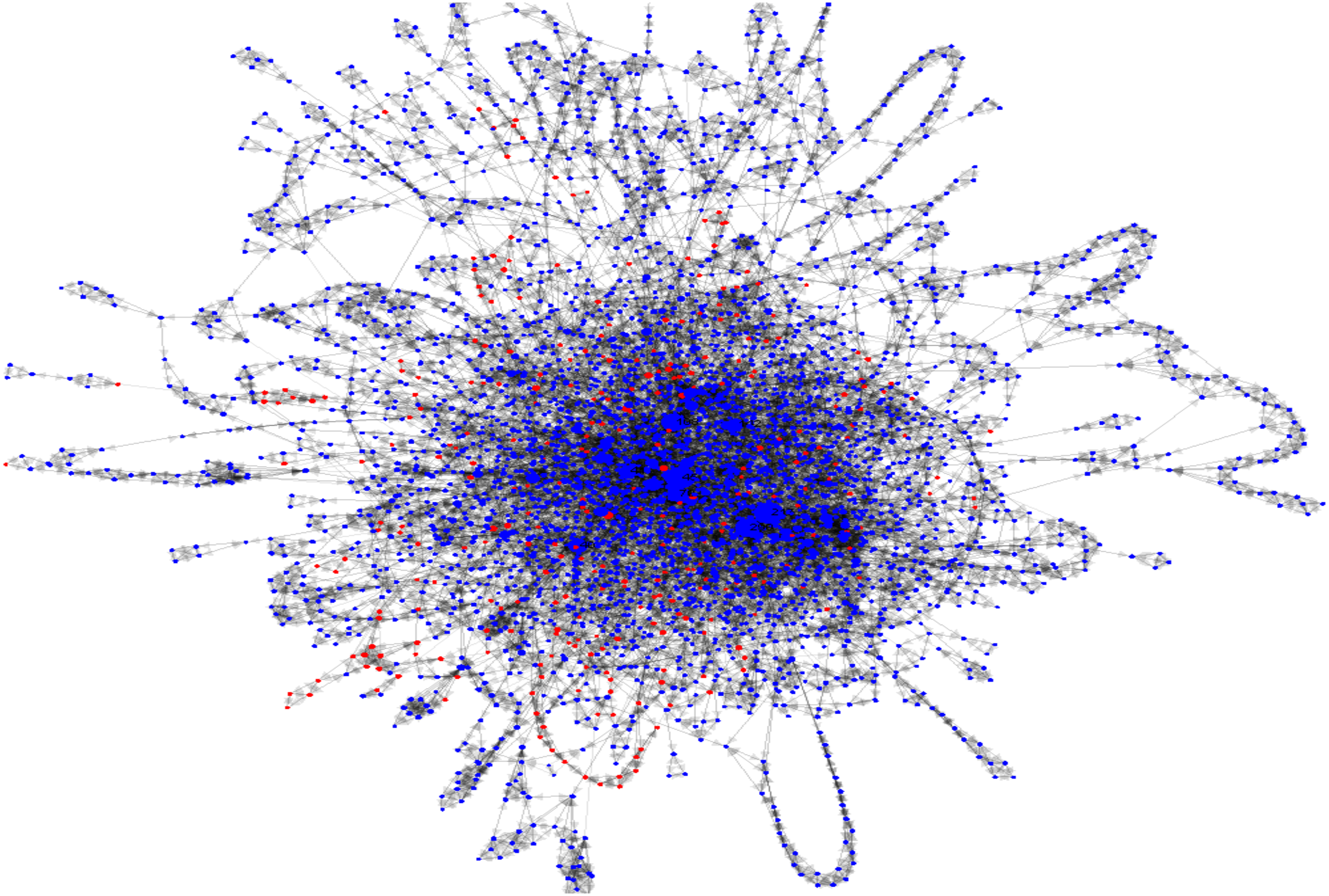
A128149: Least k such that n^k mod k = n-1.

A036236: Smallest integer k > 0 such that 2^k mod k = n

A128172: Least k such that n^k mod k = n+1.

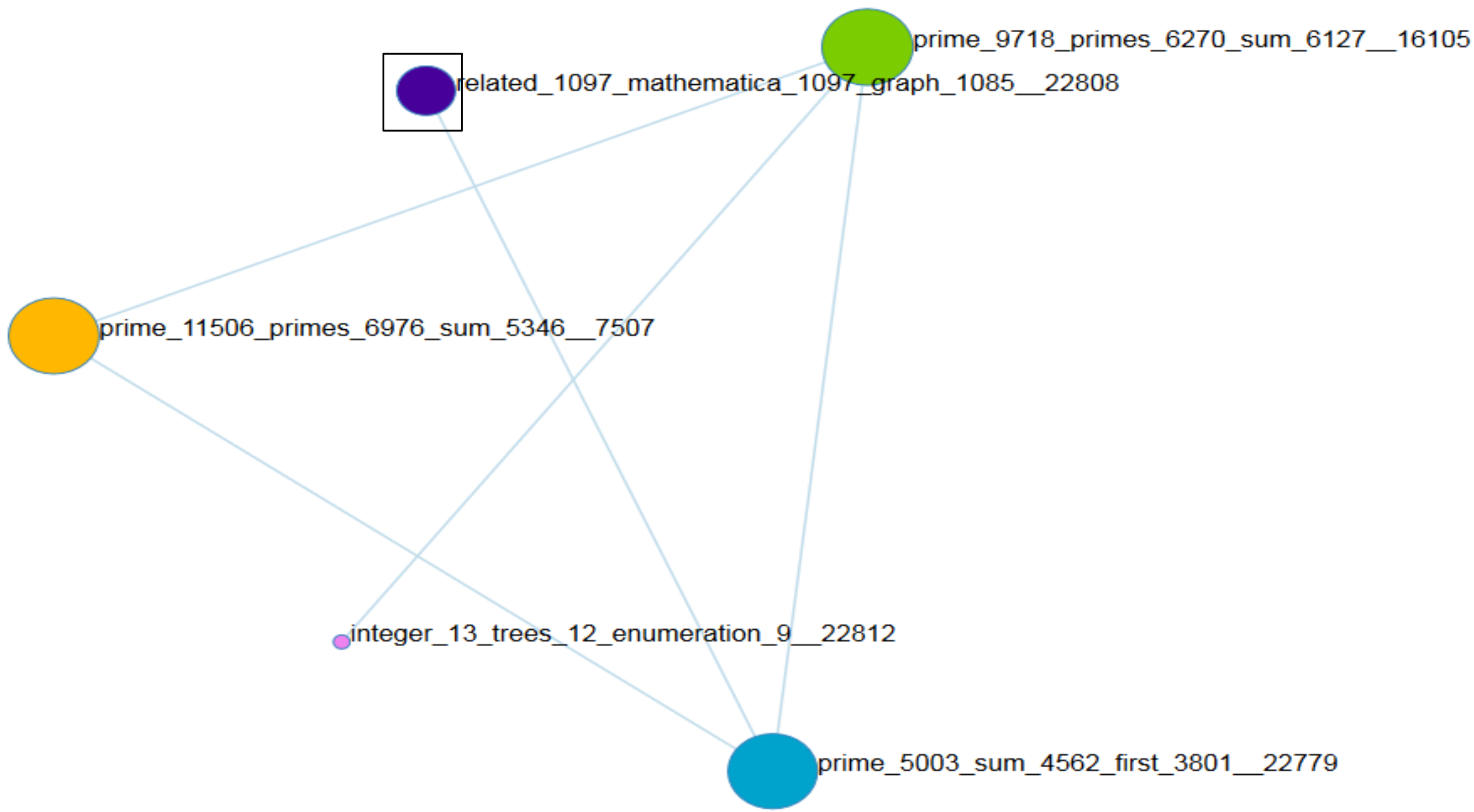


Some Graphs of OEIS



Graph of “hard” and “easy” sequences with PageRank and peeling algorithms on GraphStream.

Clustered Graph Top View



Hierarchical Clustering of the OEIS Graph

Hadley Black
Mentor: James Abello

2015 DIMACS REU

July 17, 2015

Overview

Graph Partitioning and Clustering

- *Partitioning:* Given an undirected weighted graph $G = (V, E, c, w)$ and $k > 1$, find a partition $\{V_1, ..., V_k\}$ of the vertex set V such that the total edge-cut

$$\sum_{i<j} w\Big((u,v) \in E \Big| u \in V_i, v \in V_j\Big)$$

is minimal.

Graph Partitioning and Clustering

- *Partitioning:* Given an undirected weighted graph $G = (V, E, c, w)$ and $k > 1$, find a partition $\{V_1, ..., V_k\}$ of the vertex set V such that the total edge-cut

$$\sum_{i<j} w\Big((u, v) \in E \Big| u \in V_i, v \in V_j\Big)$$

is minimal.

- *Clustering:* Partitioning, but k is not given in advance.

Overview

Outline

Bottom-up hierarchical clustering method.

Process:

1. *Pre-processing:*
   1. Vertex splitting.
   2. Graph peeling.
2. *Main steps:*
   1. Initial coarsening phase.
   2. Projection and refinement phase.
   3. Second coarsening phase with a size constraint.

Label Propagation

Label Propagation

Initialize each node as its own cluster. For each $u \in V$ (by increasing ID) move u to cluster V_i where

$$\sum_{v \in N_u} w\Big((u,v) \in E \Big| v \in V_i\Big) \text{ is maximized.}$$

Quotient Graph

Let $\{V_1, \ldots V_k\}$ be a k-clustering of G. The corresponding *quotient graph* is defined $Q = (V_Q, E_Q)$ where $V_Q = \{V_1, \ldots V_k\}$ and $E_Q = \Big\{(V_i, V_j) \Big| \exists (u,v) \in E_G, u \in V_i, v \in V_j\Big\}$. Edge and vertex weights: $w(V_i, V_j) = \sum_{u \in V_i, v \in V_j} w(u,v)$ and $w(V_i) = |V_i|$.

Projection and Refinement

For each G_i in the hierarchy (in descending order starting with G_k):

1. *Refine* the cluster at the current level using one iteration of *size-constrained* label propagation to obtain the new quotient graph G'_{i+1}.
2. *Project* the clustering to a *finer* graph in the hierarchy. For $v \in V_{G_{i-1}}$: $cluster_{G_i}[v] = cluster_{G'_{i+1}}[cluster_{G_i}[v]]$

Note: A mean cluster size constraint is applied: When applying label propagation to some vertex u, if a move to cluster V_i would result in $w(V_i) > \frac{|V|}{k}$, then that move is not considered.

Second Coarsening Phase

Build a hierarchy tree by iteratively running label propagation and computing the corresponding quotient graphs as before.

Differences:

- Initial cluster IDs for vertices in G_1 are given from refinement phase.
- As in during *refinement*, a mean cluster size constraint is applied.

Hierarchy Trees and Antichains

Hierarchy tree:

- A directed rooted tree.
- Leaves correspond to vertices in the original graph.
- Internal nodes correspond to clusters.

Hierarchy Trees and Antichains

Hierarchy tree:

- A directed rooted tree.
- Leaves correspond to vertices in the original graph.
- Internal nodes correspond to clusters.

Antichain:

- A set of vertices from the hierarchy tree such that there is no directed path between any pair.
- An antichain A is *maximal* if every *leaf* is reachable from some vertex in A.

Overview

Label Propagation Complexity

- One iteration of LP: For each vertex we examine each of its incident edges. Cost: $\mathcal{O}(n + m)$.

Label Propagation Complexity

- One iteration of LP: For each vertex we examine each of its incident edges. Cost: $\mathcal{O}(n+m)$.
- Building the quotient graph: This requires a single scan of the edge set. Cost: $\mathcal{O}(m)$.

Geometric Size Reduction

- The size of the vertex sets of graphs in the hierarchy decrease *geometrically*.
- Exponential decrease in a discrete domain.

Geometric Size Reduction

- The size of the vertex sets of graphs in the hierarchy decrease *geometrically*.
- Exponential decrease in a discrete domain.
- Thus we perform $\mathcal{O}(\log n)$ iterations of LP to obtain the full hierarchy tree.

Geometric Size Reduction

- The size of the vertex sets of graphs in the hierarchy decrease *geometrically*.
- Exponential decrease in a discrete domain.
- Thus we perform $\mathcal{O}(\log n)$ iterations of LP to obtain the full hierarchy tree.
- Complexity of the algorithm : $\mathcal{O}\big((n+m)\log n\big)$.

Overview

Added feature: Vertex Splitting

- If $deg(v) > U = \lfloor \log_2 |V| \rfloor$: v becomes $\{v_1, v_2..., v_k\}$ where $k = \lceil \frac{deg(v)}{U} \rceil$.
- For our data set $U = 18$.
- Helps avoid massive clusters centered around high degree vertices (sequences) such as the *primes* and *fibonacci numbers*.

Added feature: Peeling

- Iteratively remove vertices v where $deg(v) = 1$.
- This removes the peripheral forest of G.
- Treat each anchor point as a cluster containing the corresponding peripheral tree at the lowest level in the hierarchy.

Overview

Outline

Bottom-up hierarchical clustering method.

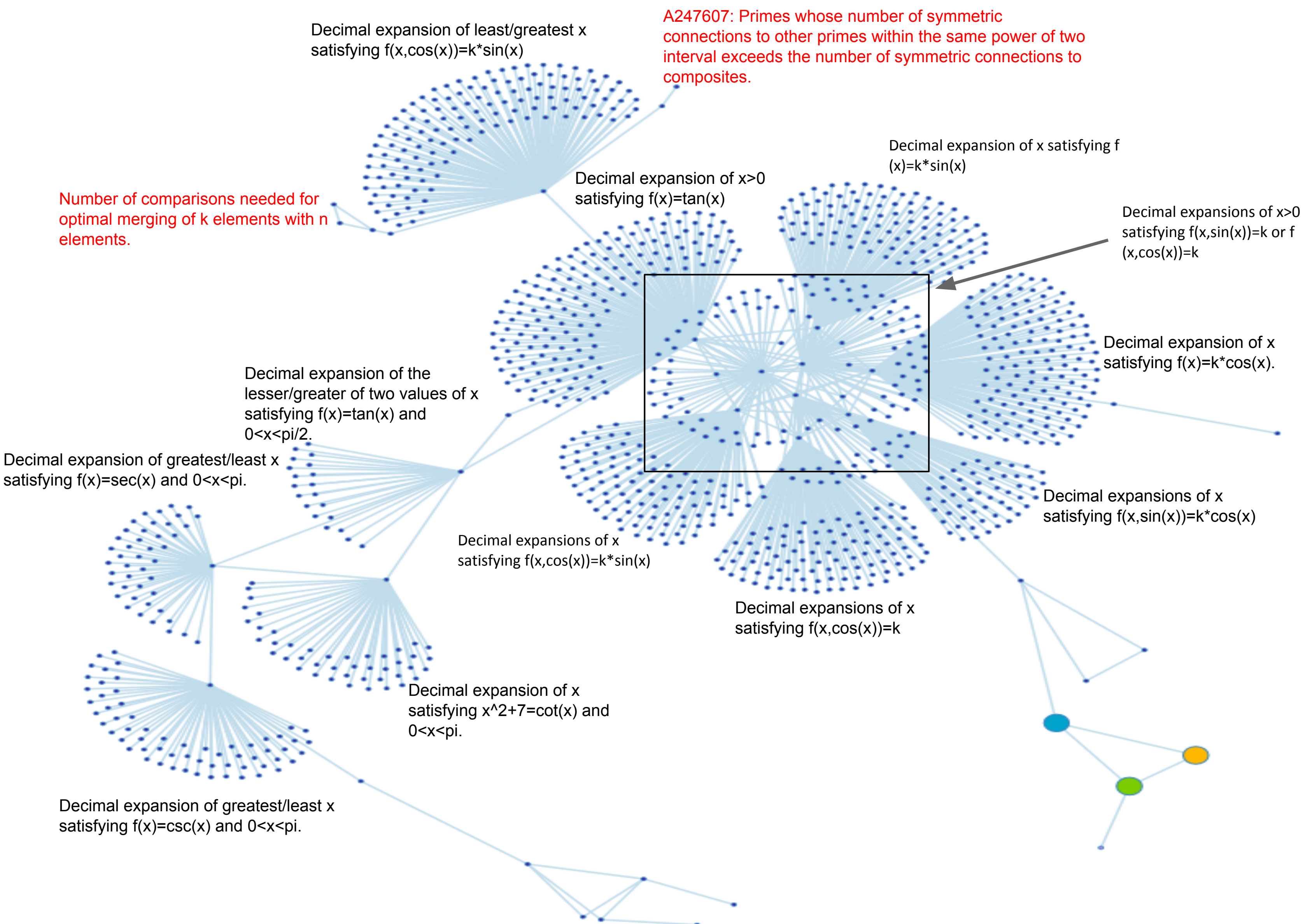
Process:

1. *Pre-processing:*
   1. Vertex splitting.
   2. Graph peeling.
2. *Main steps:*
   1. Initial coarsening phase.
   2. Projection and refinement phase.
   3. Second coarsening phase with a size constraint.

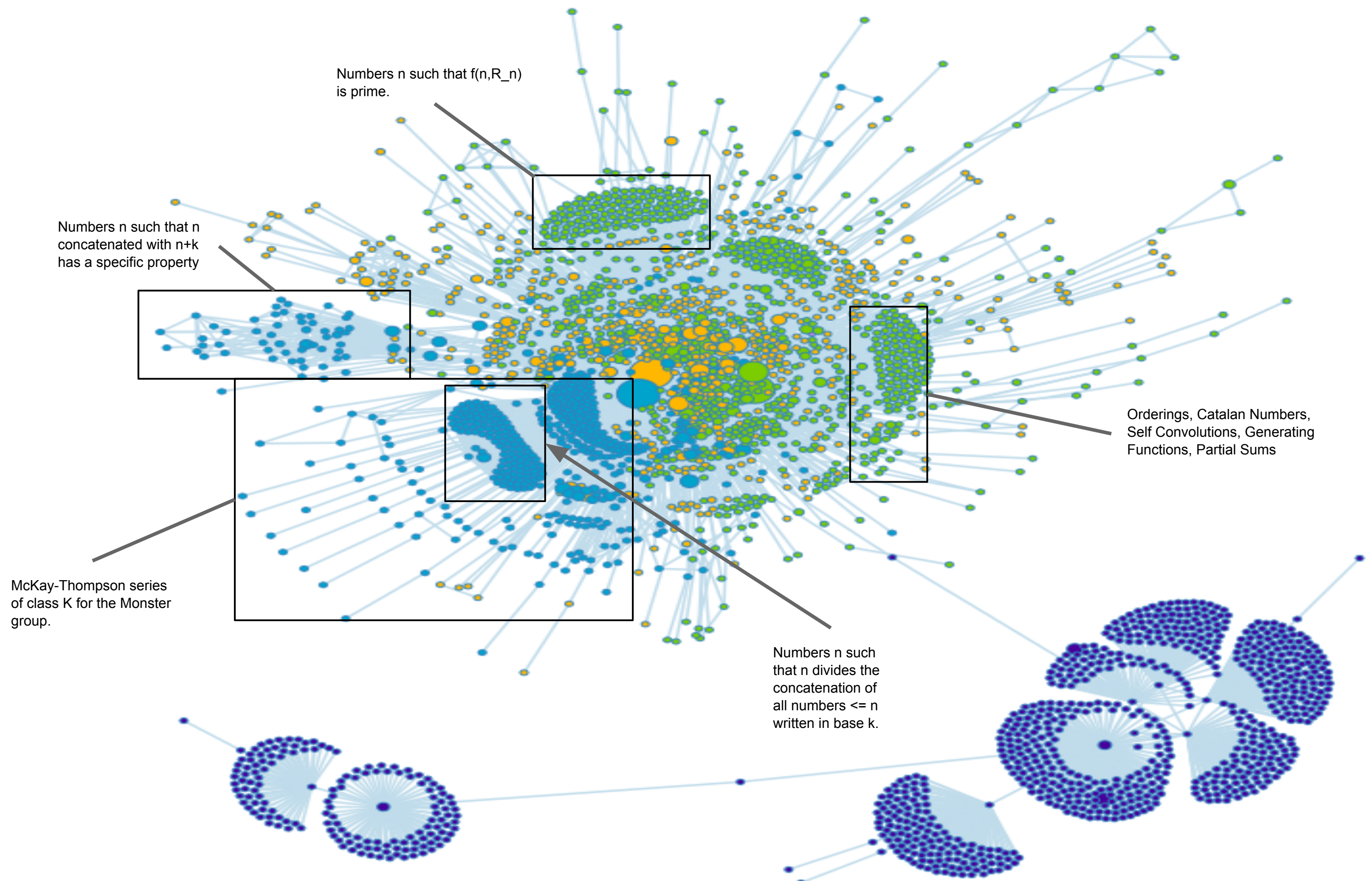
References

- (Semi-)External Algorithms for Graph Partitioning and Clustering by Akhremtsev, Sanders, and Schulz. 2014.
- Ask Graph View: The Design of a Large Scale Graph Visualization System by J. Abello, F. Van Ham, N. Krishnan. 2006.
- Simple Linear Algorithms for Mining Graph Cores by Yang Xiang. 2014.

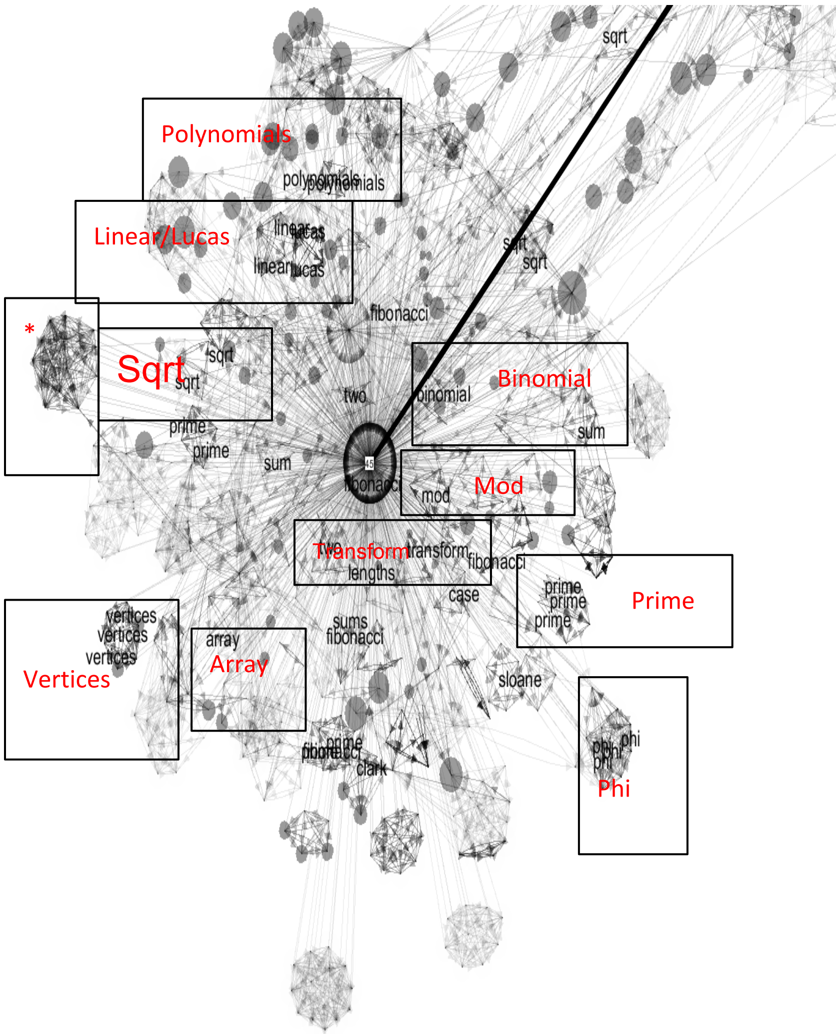
Expansion of “related”, “mathematics”, “graph” cluster



Expansion of clustered graph



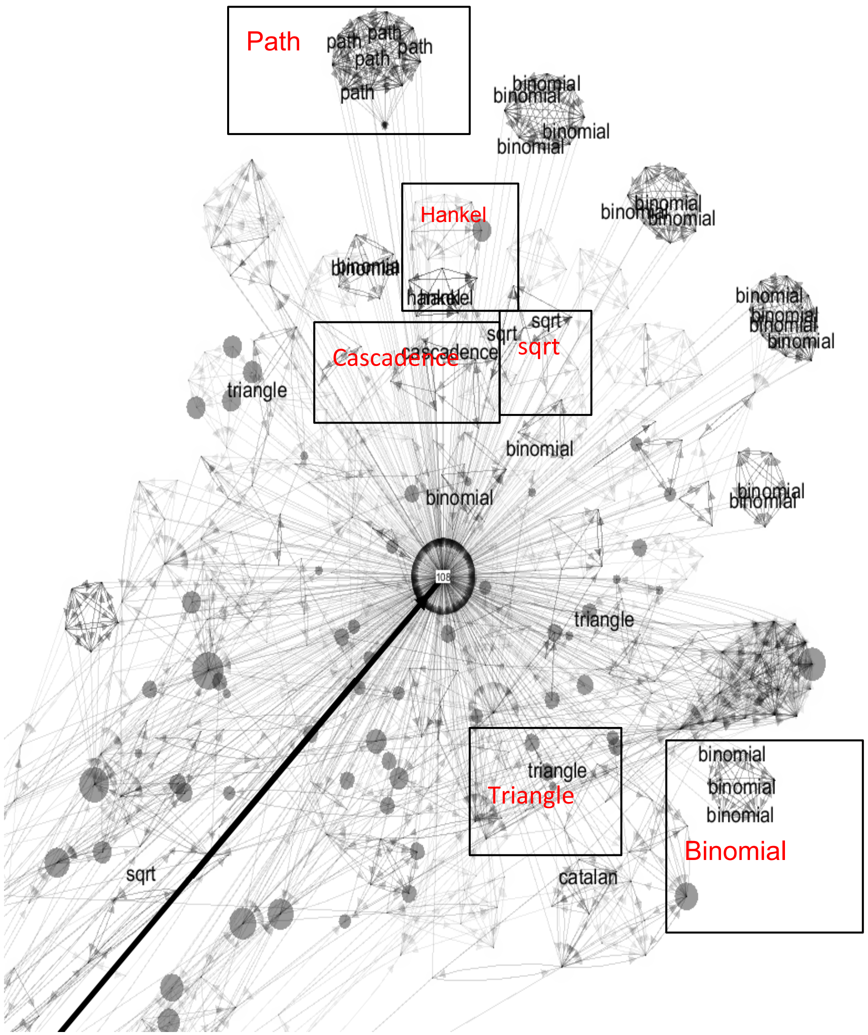
Fibonacci numbers



Observations:

- Polynomial: A57281, Coefficient triangle of polynomials (falling powers) related to Fibonacci convolutions.
- Linear: A82411: Second-order linear recurrence sequence with a(0) and a(1) coprime that does not contain any primes.
- Lucas: A115313, GCD(Lucas(n)+1,Fibonacci(n)+1).
- *Digital sum analogue (in base k) of the Fibonacci recurrence.
- Sqrt: A6190, Denominators of continued fraction convergents to (3+Sqrt[13])/2.
- Binomial: a(n)= A99572, sum{k=0..n, F(n-k+1)*binomial(k/2+3, 3)(1+(-1)^k)/2}.
- Mod: A51834, Fibonacci(Pn-1) mod Pn, where Pn is the n-th prime.
- Transform: A99484, A Chebyshev transform of the sequence 1,1,3,9,27 with g.f. (1-2x)/(1-3x).
- Prime: A215818, Fibonacci numbers F(n) that have no prime factor congruent to 5 or 7 mod 8 to an odd power.
- Vertices: A182465, Number of vertices into building blocks of 3d objects with 7 vertices.
- Array: A243613, Irregular triangular array of denominators of the positive rational numbers.
- Phi: A214887, Let phi = 1/2*(1 + sqrt(5)) denote the golden ratio A001622. This sequence is the simple continued fraction expansion of the constant c := 6*sum {n = 1..inf} 1/7^floor(n*phi) (= 36*sum {n = 1..inf} floor(n/phi)/7^n) = 0.87718 67194 00499 51922 ... = 1/(1 + 1/(7 + 1/(7 + 1/(49 + 1/(343 + 1/(16807 + 1/(5764801 + ...))))))).

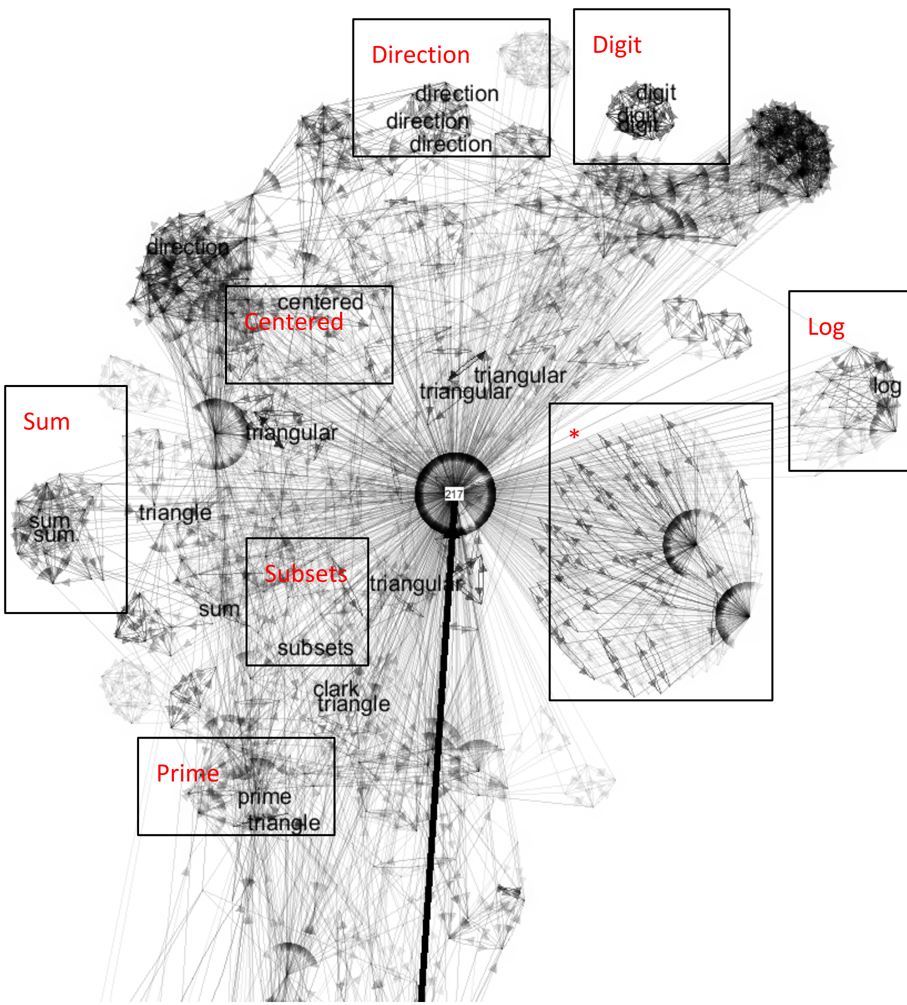
Catalan numbers



Observations:

- Path: A258181, Sum over all Dyck paths of semilength n of products over all peaks p of 2^(x_p-y_p), where x_p and y_p are the coordinates of peak p.
- Hankel: A156362, Hankel transform is 7^C(n+1,2),
- Cascadence: A120914, Cascadence of (1+2x)^2; a triangle, read by rows of 2n+1 terms.
- Sqrt: A68767, G.f.: (1-sqrt(1-20*x*(1-4*x)))/(10*x).
- Triangle: A125275, Eigensequence of triangle A039599.
- Binomial: A234464, 5*binomial(8*n+5, n)/(8*n+5).

Triangular numbers



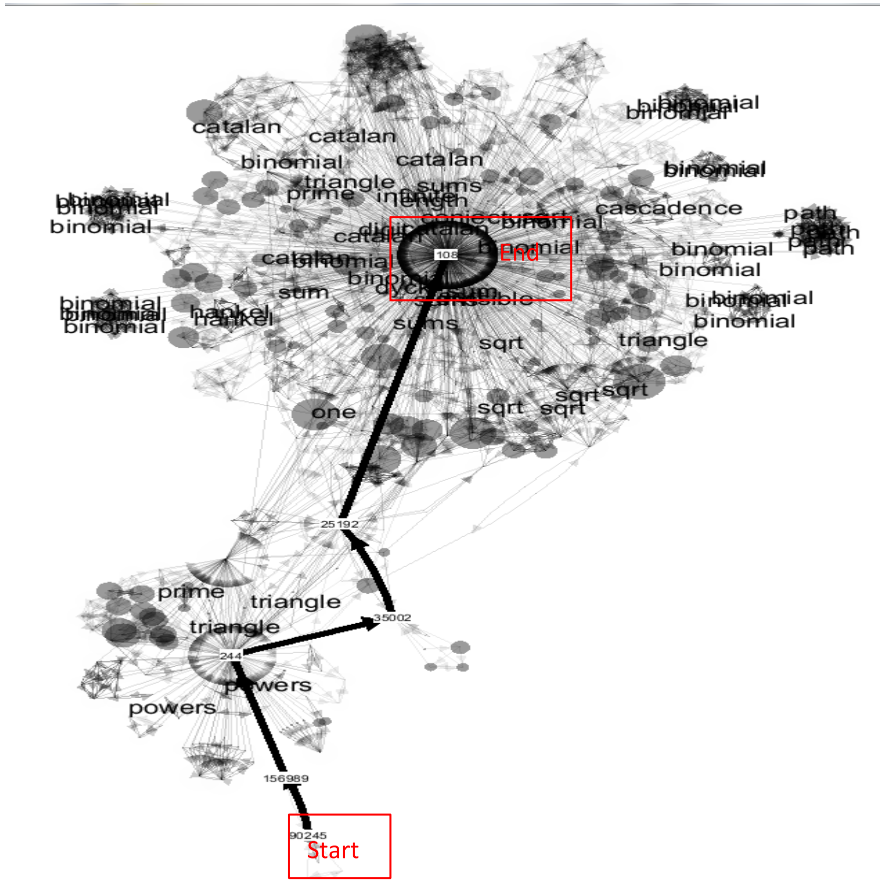
Observations:

- Direction: A139595, Sequence found by reading the line from 0, in the direction 0, 5,... and the same line from 0, in the direction 0, 13,..., in the square spiral whose vertices are the triangular numbers.
- Digit: A241792, Triangular numbers which have one or more occurrences of exactly nine different digits.
- Centered: A254627, Indices of centered pentagonal numbers that are also triangular numbers.
- Log: A132033, Product{0<=k<=floor(log_9(n)), floor (n/9^k)}, n>=1.
- Sum: A249120, Conjecture: gives an identity for the sum of all divisors of all positive integers <= n.
- *Triangular numbers composed of digits {x,y,z}.
- Subset: A248141, Table read by rows: n-th row contains all subsets of consecutive numbers of 1..n.
- Prime: A228430, Number of ways to write n = x + y (x, y > 0) with x^4 + y*(y+1)/2 prime.

A90245 - A108

Source: Maximum numbers of cards that would have no SET in an n-attribute version of the SET card game

Destination: Catalan numbers



Observations:

- Connects a scientist (James Abello) to a sequence.

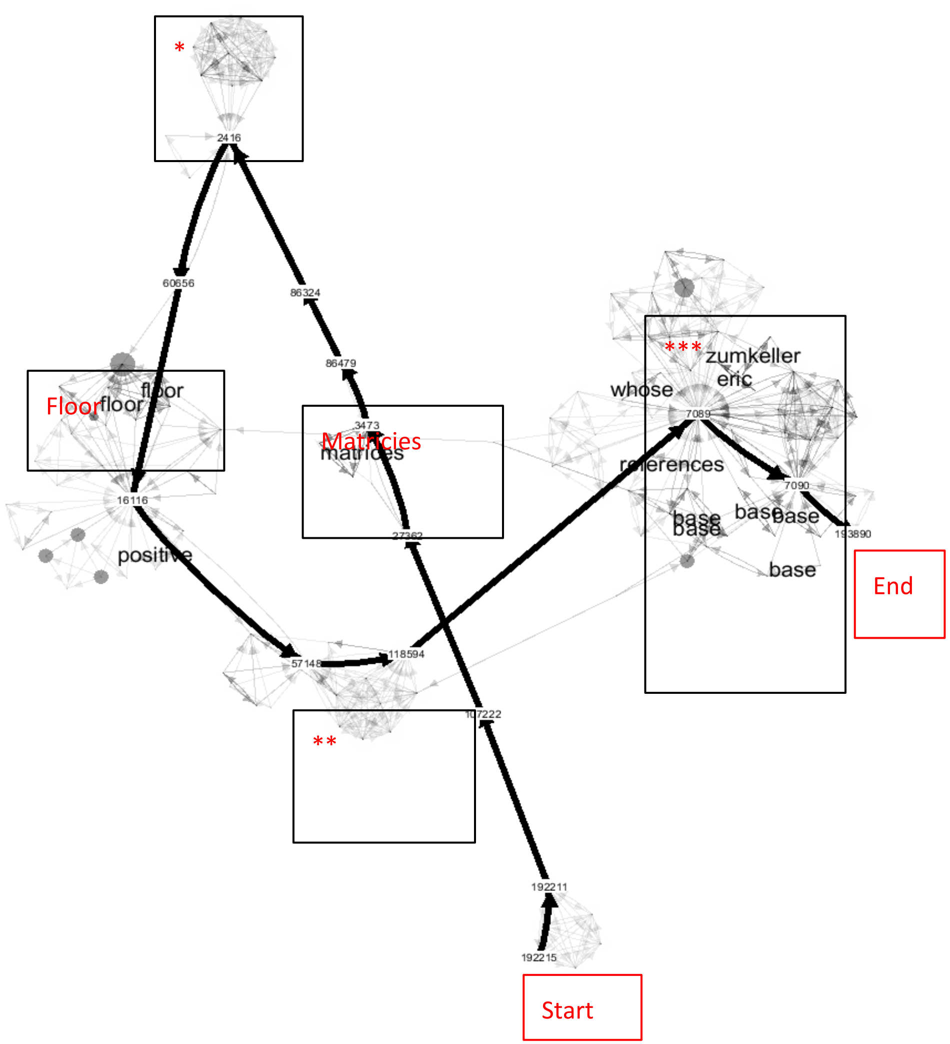
Statistics (before/after):

- Nodes: 243/1231
- Edges: 1548/4185
- Displayed Keywords: 281/18

A192215 - A193890

Source: Number of zero trace primitive elements in Galois field GF(11^n).

Destination: Primes p such that replacing any single decimal digit d with 3*d produces another prime (obviously p can contain only digits 0, 1, 2 or 3).



Observations:

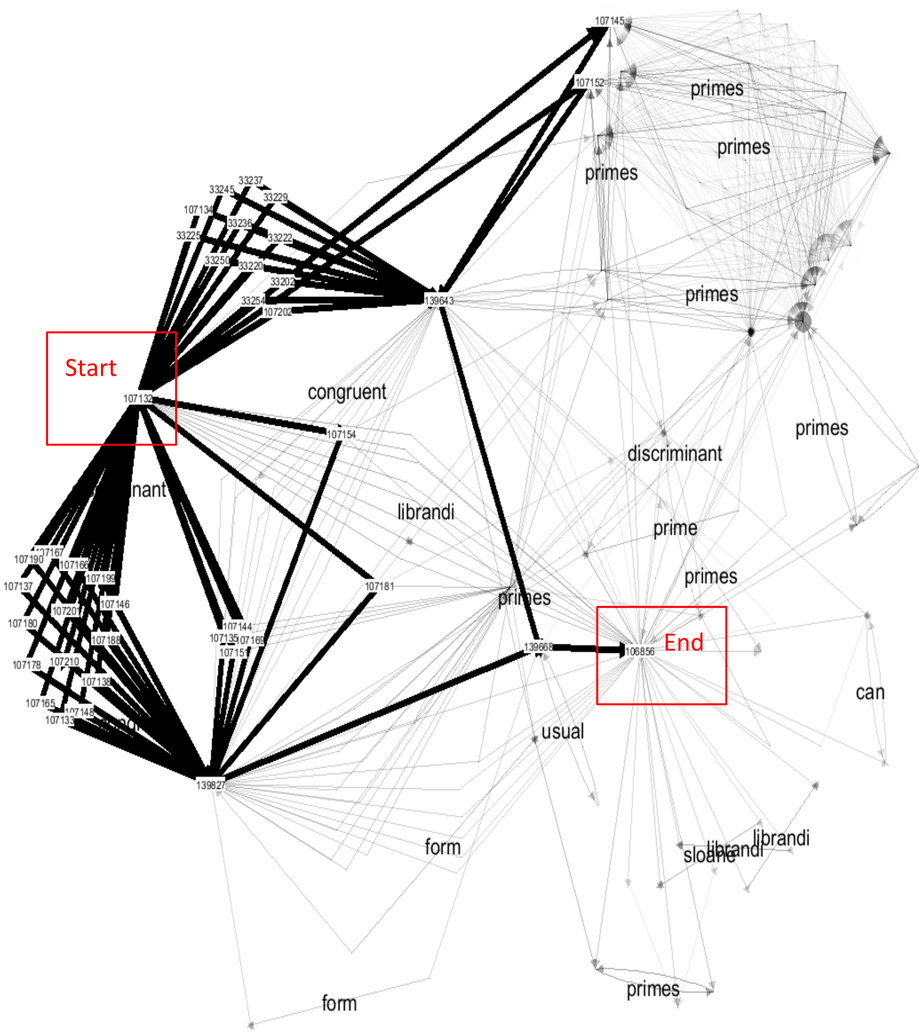
- Source and Destination have keyword “hard” and “more”.
- Source and Destination both consist of only 5 terms.
- The shortest path is of length 14.

- Matricies: A3473, a(n) is the number of n X n circulant invertible matrices over GF(2).
- *Number of binary arrangements without adjacent 1's on n X n objects.
- Floor: A133629, a(n)=p^floor(n/2)+(p^floor((n+1)/2)-p)/(p-1) where p=5.
- **Palindromes in base k.
- ***Base k representations of n with special properties or operations on them.

A107132 - A106856

<u>Source:</u> Primes of the form 2x^2+13y^2. T.D. Noe, 2005.

<u>Destination:</u> Primes of the form x^2+xy+2y^2, with x and y nonnegative. T.D. Noe, 2005.

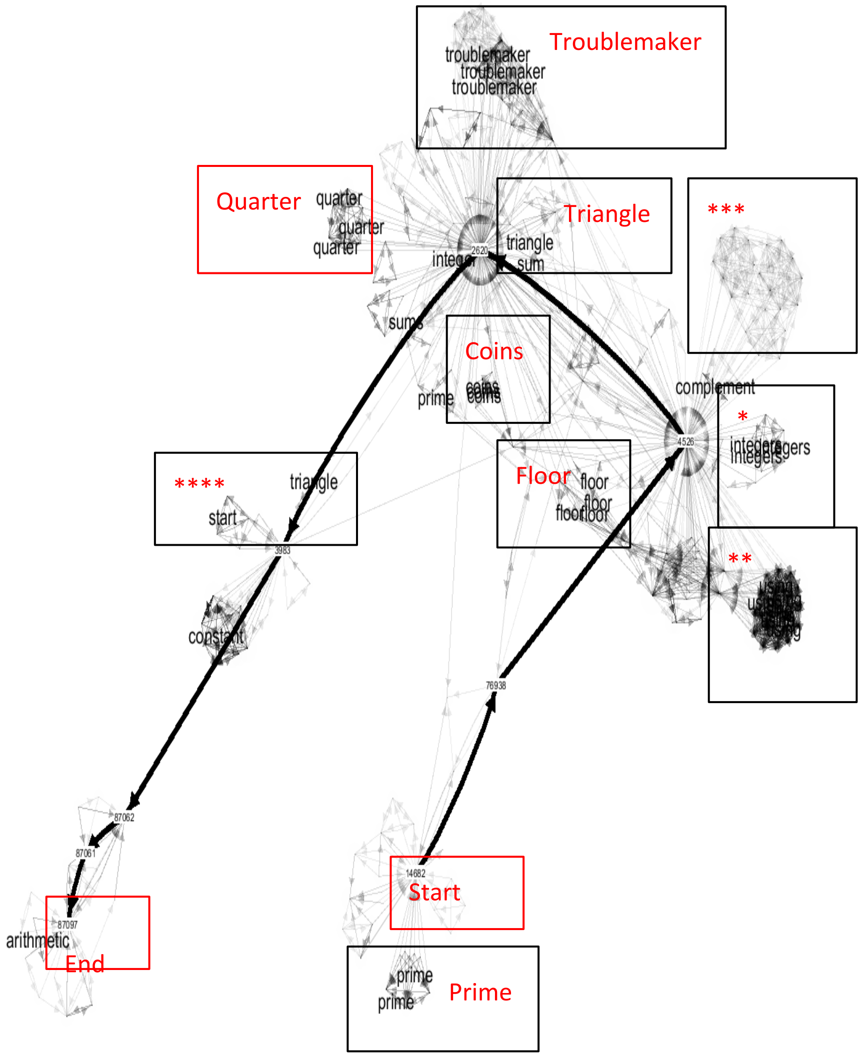


<u>Observations:</u>

- The source and destinations are the top two sequences with the most cross references.
- There are 34 paths of length 4 from the source to the destination.
  - [redacted]
- Sparse considering the fact that both sequences involve prime numbers.
- Generally primes of the form ax^2+bxy+cy^2.

A014682 - A087097

Source: The Collatz or 3x+1 function.
Destination: Lunar Primes.



Observations:

- One shortest path of length 7.
- Prime: A214892, Similar to the Fibonacci recursion starting with (4, 1), but each new non-prime term is divided by its least prime factor.
- **Floor(a*n/b).
- Floor: A131242, a(n)=(1/2)*floor(n/10)*(2n-8-10*floor(n/10)).
- *Number of necklaces with k black beads and n-k white beads.
- ***Sequences involving the Kaprekar map.
- Coins: A230548, Twin hearts patterns packing into nXn coins.
- Triangle: Number of nondegenerate triangles that can be made from rods of length 1,2,3,4,...,n.
- Troublemaker: Apart from the initial term this is the elliptic troublemaker sequence R_n(1,5) (also sequence R_n(4,5)) in the notation of Stange (see Table 1, p.16).
- Quarter: A257023, Number of terms in the quarter-sum representation of n.
- ****:Triangle whose rows are sequences of increasing and decreasing square,cubes,etc.

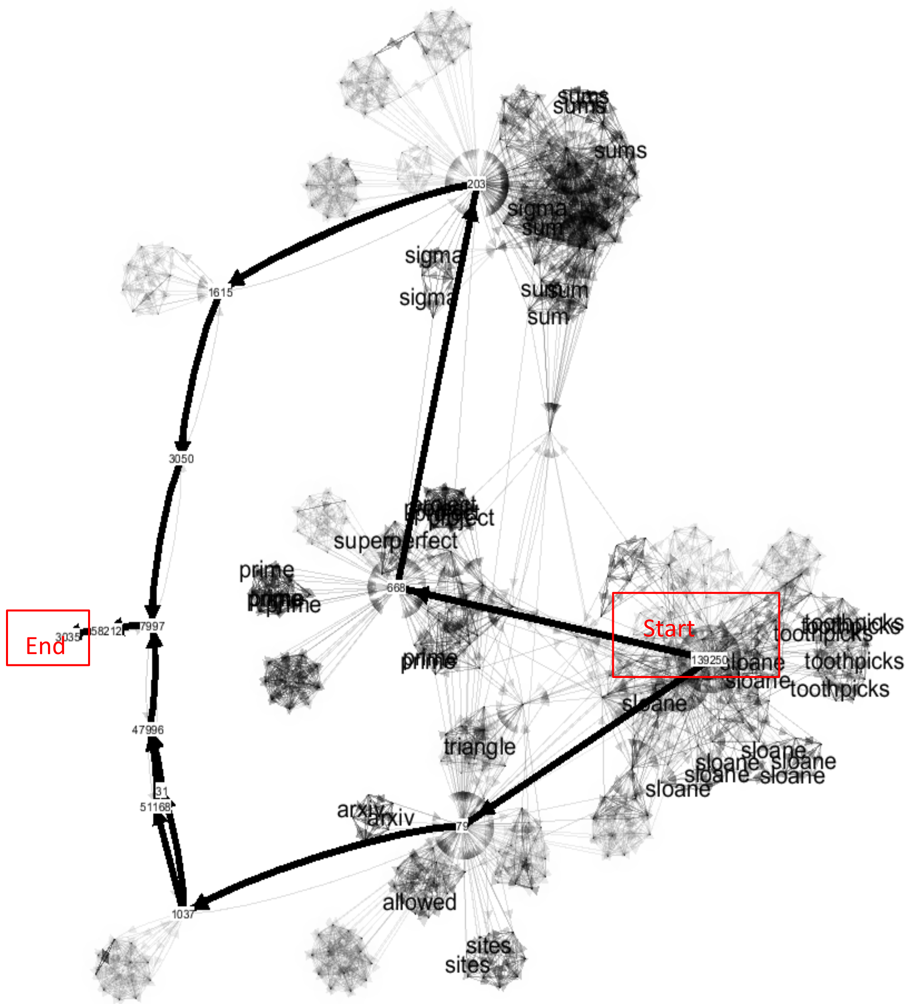
Statistics:

- Nodes: 350
- Edges: 1532
- Displayed Keywords: 21

A139250 - A3035

<u>Source:</u> Toothpick sequence. O. Pol, 2008.

<u>Destination:</u> Maximal number of 3-tree rows in n-tree orchard problem. N.J.A. Sloane.



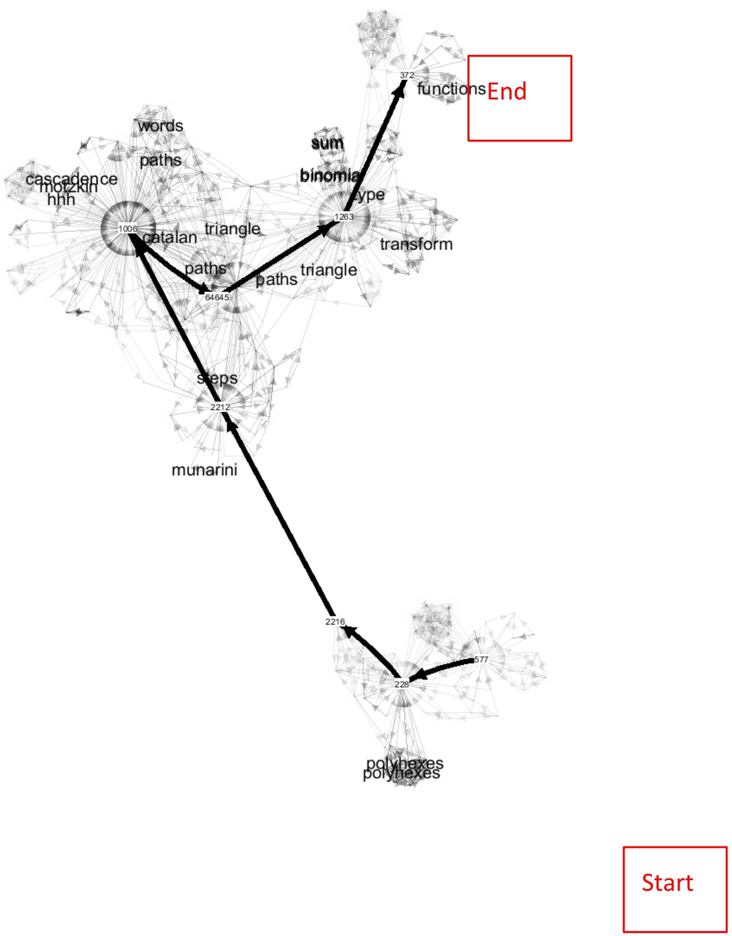
<u>Observations:</u>

- There exist three shortest paths of length 7.
  - A668: Mersenne primes (of form 2^p - 1 where p is a prime).
  - A203: a(n) = sigma(n), the sum of the divisors of n.
  - A1615: Dedekind psi function: n * Product_{p|n, p prime} (1 + 1/p).
  - A3050: Number of primitive sublattices of index n in hexagonal lattice.
  - A7997: a(n) = ceiling((n-3)(n-4)/6).
  - A58212: 1 + floor(n*(n-3)/6).

<u>Statistics:</u>

- Nodes: 396
- Edges: 3442
- Displayed Keywords: 12

A577 - A372



<u>Source:</u> Number of triangular polyominoes with n cells. N.J.A. Sloane, 1973.

<u>Destination:</u> Dedekind numbers. N.J.A. Sloane, 1973.

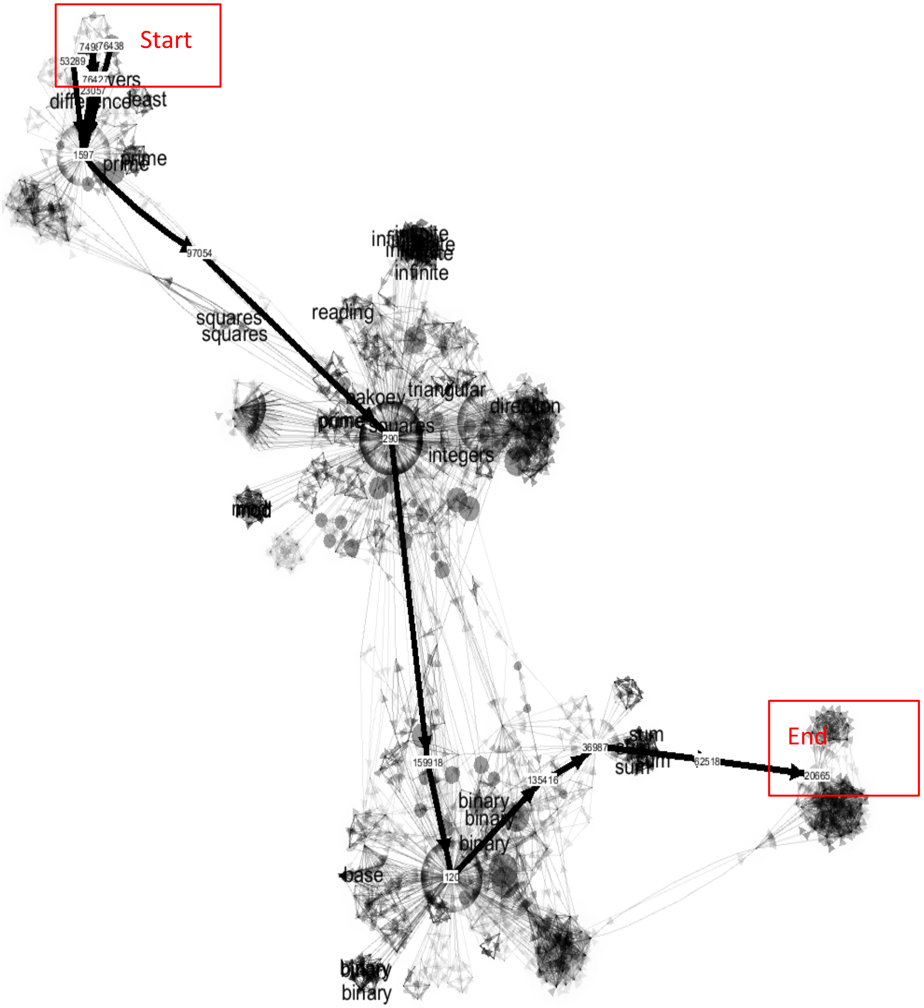
<u>Observations:</u>

- Both sequences are labeled “hard” and “nice” .

- There exist only one shortest path of length 7.
    - A228: Number of hexagonal polyominoes (or planar polyhexes) with n cells.
    - A2216: Restricted hexagonal polyominoes (cata-polyhexes) with n cells.
    - A2212: Number of restricted hexagonal polyominoes with n cells.
    - A1006: Motzkin numbers: number of ways of drawing any number of nonintersecting chords joining n (labeled) points on a circle.
    - A64645: Table where the entry (n,k) (n >= 0,k >= 0) gives number of Motzkin paths of the length n with the minimum peak width of k.
    - A1263: Catalan triangle.

- Nodes: 395.
- Edges: 1459.
- Displayed Keywords: 15.

A74981 - A20665



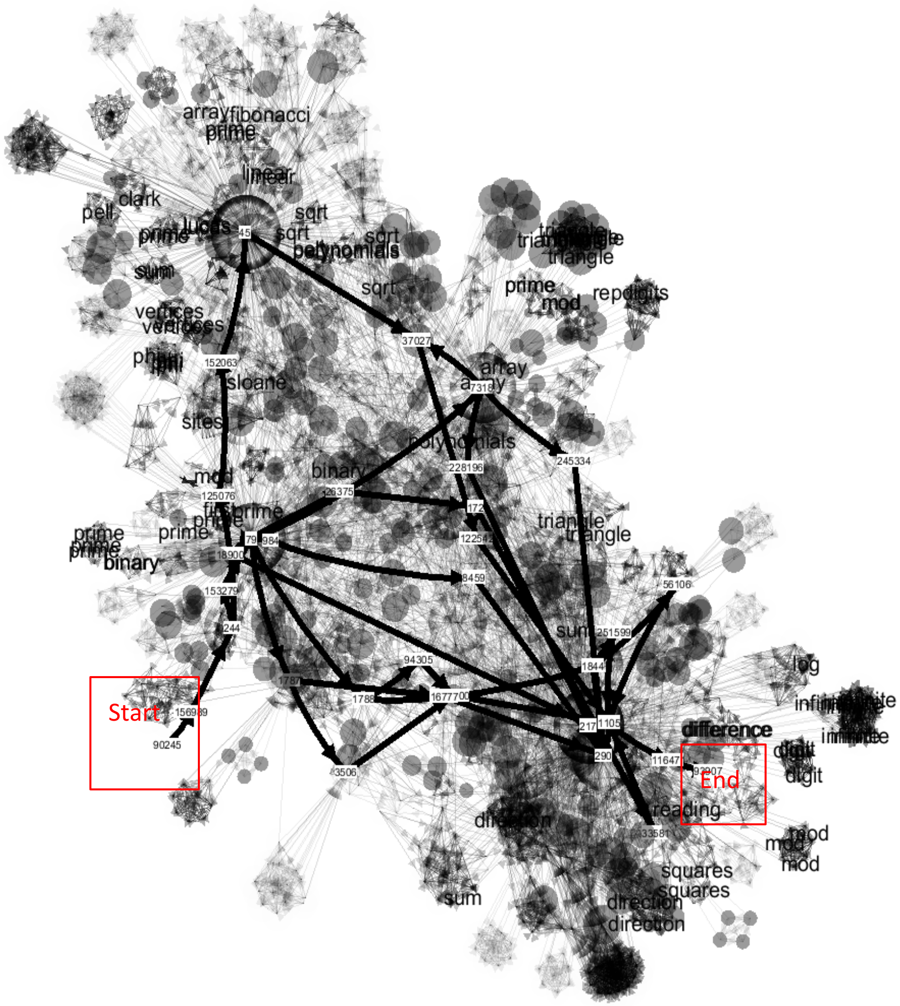
Source: Conjectured list of positive numbers which are not of the form r^i-s^j, where r,s,i,j are integers with r>0, s>0, i>1, j>1.

Destination: Conjectured maximal exponent k such that n^k does not contain a digit zero in its decimal expansion.

Observations:

- Both sequences are conjectured.
- There are 4 shortest path of length 10.
- All shortest paths pass through the following nodes.
  - A1597: Perfect powers: m^k where m > 0 and k >= 2.
  - A97054: Nonsquare perfect powers.
  - A290: The squares.
  - A159918: Number of ones in binary representation of n^2.
  - A120: Number of 1's in binary expansion of n.
  - A135416: A036987(n)*(n+1)/2.
  - A36987: Fredholm-Rueppel sequence.
  - A62518: Conjectural largest exponent k such that n^k does not possess all of the digits 0 through 9 (in decimal notation) or 0 if no such k exists (if n is a power of 10).
- Nodes: 730.
- Edges: 4213.
- Displayed Keywords: 15.

A90245 - A93907



<u>Source:</u> Maximum numbers of cards that would have no SET in an n-attribute version of the SET card game.

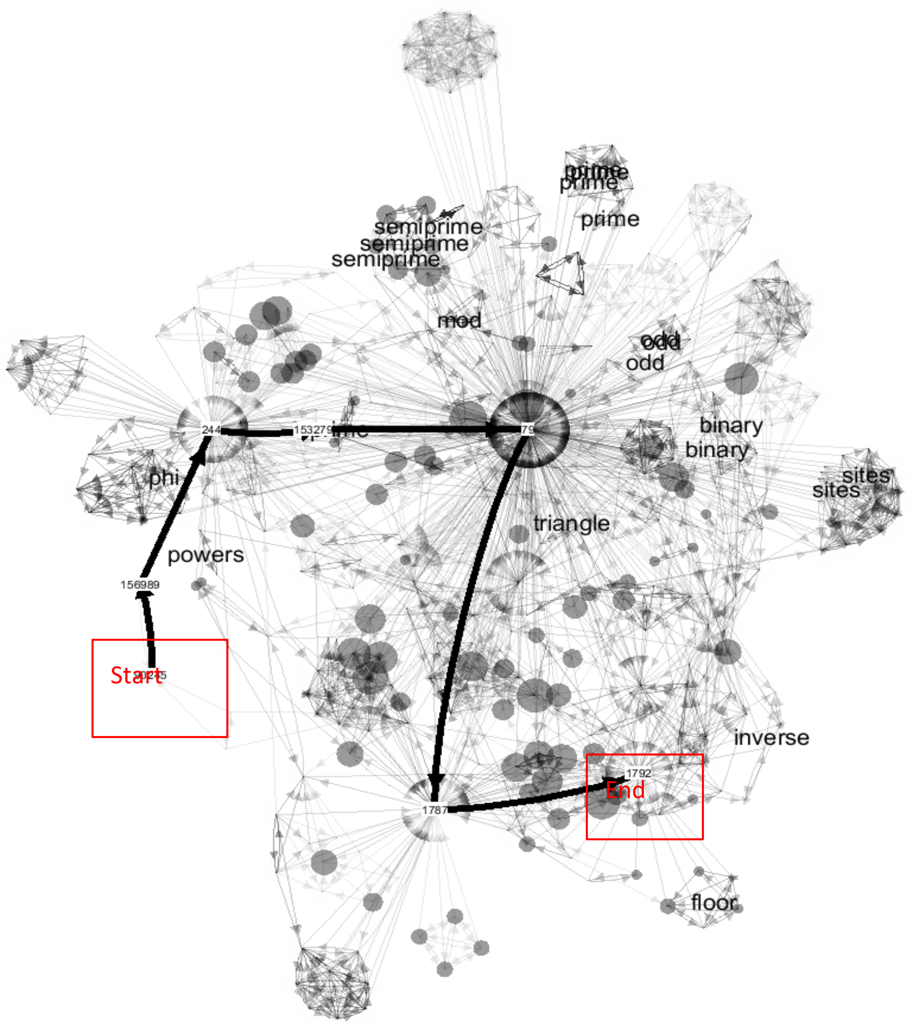
<u>Destination:</u> Number of elements in the n-th period of the periodic table.

<u>Observations:</u>

- Connects a scientist to a new field.
- There exist 13 shortest paths of length 10.

- Nodes: 1569.
- Edges: 10769.
- Displayed Keywords: 26.

A90245 - A1792



Source: Maximum numbers of cards that would have no SET in an n-attribute version of the SET card game. Hans Havermann, 2004.

Destination: a(n) = (n+2)*2^(n-1). N.J.A. Sloane.

Observations:

- Connects contributor to contributor (James Abello to David Molnar).

- There exist one shortest path of length 6.
  - A156989: Largest size of a subset of {1,2,3}^n that does not contain any combinatorial lines (i.e. strings formed by 1, 2, 3, and at least one instance of a wildcard x, with x then substituted for 1, 2, or 3, e.g. 12x3x gives the combinatorial line 12131, 12232, 12333.). T. Tao, 2009.
  - A244: Powers of 3. N.J.A. Sloane.
  - A153279: Eigentriangle by rows, T(n,k) = A000079(n-k) * (diagonalized matrix of (1,1,3,9,27,81,...)). G.W. Adamson, 2008.
  - A79: Powers of 2: a(n) = 2^n. N.J.A. Sloane, 1991.
  - A1878: n*2^(n-1). N.J.A Sloane.

- Nodes: 503.
- Edges: 2483.
- Displayed Keywords: 11.

Multiple Shortest Paths

- Idea: Given a tree and a vertex not on the tree, find the shortest path from the vertex to any node on the tree
- More computationally efficient than initial idea

Flow: Easies to Hards

- Idea: Attach an artificial source and sink to all easies and all hards resp., then find maximum flow from source to sink
- Tells us all easies that reference hards

Outline

1. Technologies
   - Graph Visualization
   - Web Programming
2. Application
   - Local Development
   - Database
   - Web Application
3. Stumbling Blocks and Continuing Work

Outline

1. Technologies
   - Graph Visualization
   - Web Programming
2. Application
   - Local Development
   - Database
   - Web Application
3. Stumbling Blocks and Continuing Work

The GraphStream Library

A Java Library for Visualizaiton of Static and Dynamic Graphs

- Platform used for prior work (Twitter Map, Protein Evolution)
- Produces Force-directed Node-Link layout

Architecture

- Core - nodes, edges, attributes
- Algo - common graph algorithms (BFS, Dijkstra, PageRank)
- UI - display based on attributes and force directed layout

D3
Data Driven Documents

- JavaScript library with wide array of graphical capabilities
- Most important one for us is the Force Directed Layout
- Interacts with HTML DOM
- Event-driven
- Uses CSS styling

How the Internet Works

- Browser HTTP GET request to Web Server
- Response comes as XML document specifying content and layout
- Browser renders the result according to HTML specifications
- Ordinary users never see HTTP headers or worse, transport protocols
- HTML defines layout, CSS style, JavaScript behavior

```
▶<div class="mainmenu" id="mainmenu">…</div>
▶<div class="controls" id="controls">…</div>
 <script src="assets/js/fdl.js"></script>
▼<svg viewBox="0 0 1280 612">
  ▼<g class="fullField">
     <rect width="1280" height="612"></rect>
    ▼<g class="zoomableField">
      ▶<g class="links">…</g>
      ▼<g class="nodes">
        ▶<g class="node" transform="translate(631.4739175130363,259.68930531504554)">…</g>
        ▶<g class="node" transform="translate(558.454164499278,358.79740475424046)">…</g>
        ▶<g class="node" transform="translate(548.5770762865628,315.77809568661456)">…</g>
        ▶<g class="node" transform="translate(712.0695219416174,287.95113604940417)">…</g>
        ▶<g class="node" transform="translate(552.5239162285058,268.7816143557891)">…</g>
```

```css
.node circle {
  cursor: pointer;
  stroke: #3182bd;
  pointer-events: all;
}

.node text {
  font-family: sans-serif;
  pointer-events: none;
  text-anchor: start;
  background-color:white;
}
```

```
    nodes = flatten(root);
    mergeInLinks(addLinks);

    link_elements = link_elements.data(links, function(d) {
        return d.id;
    });

    link_elements.exit().remove();

    link_elements.enter().insert("line", ".node").attr("class", "link")
            .attr("stroke-width", strokeWidth);

    force.nodes(nodes).links(links).start();
}
```

Basic Web Service Types

- Simple Object Access Protocol (SOAP)
- REpresentative State Transfer (REST)
- SOAP is language, platform, and transport protocol independent
- Well-standardized, built in error-handling, but XML-only and steep learning curve
- REST addresses resources by URL and can receive varying response types
- Younger standard, gain efficiency by ditching XML, HTTP only, shorter learning curve

Jax-RS (Jersey)

- Java library for creating RESTful Web Services
- Runs on a web server in a .war file (instead of .jar)
- Specifies resources, access, response, and concurrency types as Java annotations to methods
- Allows for String or Streaming responses

```java
@GET
@Path("{newNode}/{pathTo}/{includeNeighborhoods}")
@Produces(MediaType.APPLICATION_JSON)
public String getAddition(@PathParam("newNode") String newNode,
        @PathParam("pathTo") String existing, @PathParam("include
    try {
```

TomEE and Maven

- Apache projects
- TomEE is a version of Tomcat for JavaEE
- Maven build tool resolves dependencies and executes builds
- Well-documented and popular tools with good compatibility

```
        <dependency>
            <groupId>com.google.guava</groupId>
            <artifactId>guava</artifactId>
            <version>18.0</version>
        </dependency>
        <dependency>
            <groupId>mysql</groupId>
            <artifactId>mysql-connector-java</artifactId>
            <version>5.1.35</version>
        </dependency>
    </dependencies>

    <build>
        <finalName>ROOT</finalName>
        <plugins>
            <plugin>
                <groupId>org.apache.openejb.maven</groupId>
                <artifactId>tomee-maven-plugin</artifactId>
                <version>${tomee.version}</version>
                <configuration>
                    <tomeeVersion>${tomee.version}</tomeeVersion>
                    <tomeeClassifier>jaxrs</tomeeClassifier>
                </configuration>
            </plugin>
            <plugin>
                <groupId>org.apache.maven.plugins</groupId>
                <artifactId>maven-war-plugin</artifactId>
                <version>2.0</version>
                <configuration>
                    <archive>
                        <manifest>
                            <addClasspath>true</addClasspath>
                            <classpathPrefix>lib/</classpathPrefix>
                        </manifest>
                    </archive>
```

Outline

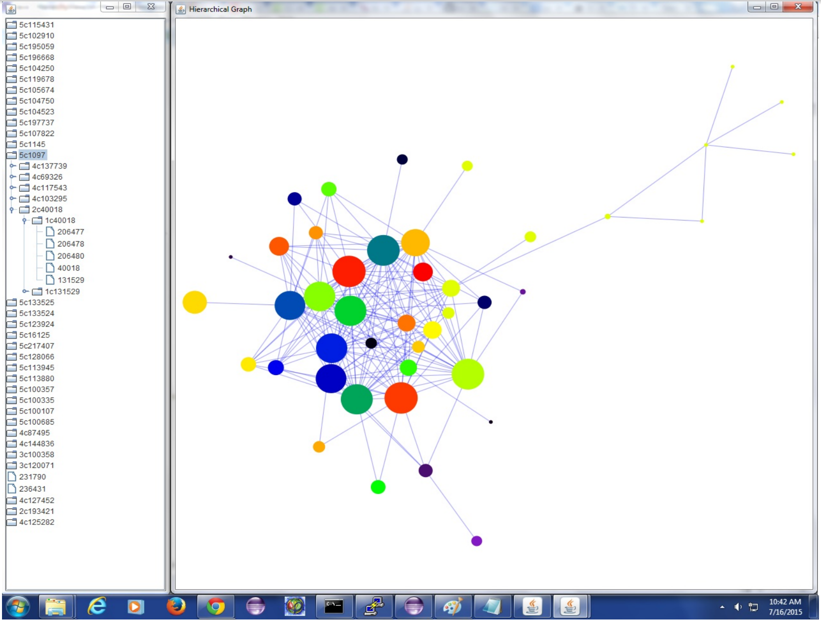
1. Technologies
   - Graph Visualization
   - Web Programming
2. Application
   - Local Development
   - Database
   - Web Application
3. Stumbling Blocks and Continuing Work

Java Application(s) using GraphStream

- Gaining initial understanding of graph structure visually
- Word Co-Occurrence Graph (Dynamic)
- Shortest Path
- Twitter Map
- Prototyping of labeling and sizing

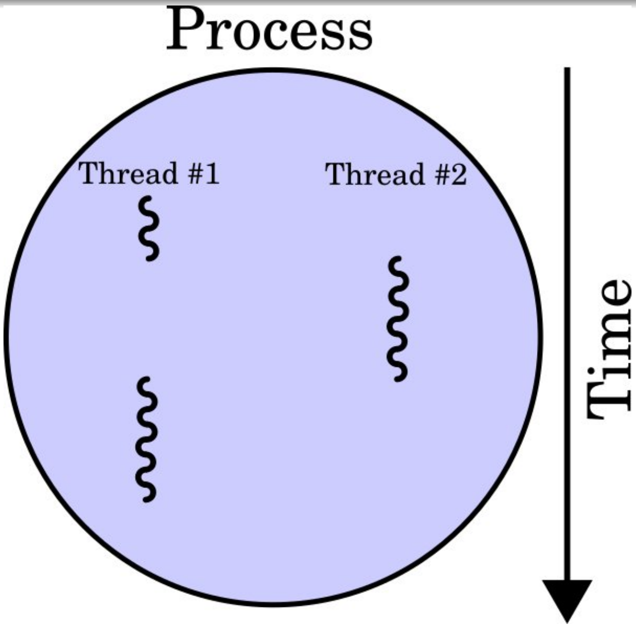
Viewing Hierarchies

- Created HierarchyViewer Application
- Reads Hierarchy Tree from recursively-defined structure
- Upon clicking a node, expands into its children, recomputes part of edge set
- Used JTree (folder layout) to show full hierarchy and allow collapsing of nodes
- Still in use as a convenient prototyping tool for clustering


5c115431
5c102910
5c195059
5c196668
5c104250
5c119678
5c105674
5c104750
5c104523
5c197737
5c107822
5c1145
5c1097
4c137739
4c69326
4c117543
4c103295
2c40018
1c40018
206477
206478
206480
40018
131529
1c131529
5c133525
5c133524
5c123924
5c16125
5c217407
5c128066
5c113945
5c113880
5c100357
5c100335
5c100107
5c100685
4c87495
4c144836
3c100358
3c120071
231790
236431
4c127452
2c193421
4c125282
Hierarchical Graph
10:42 AM
7/16/2015

GraphStream Limitations

- Threading model requires graph interaction to be done in a single thread
- Threads may be main or event dispatch only, even for Core
- Decidedly inconvenient for a user of OEIS to download and run full application

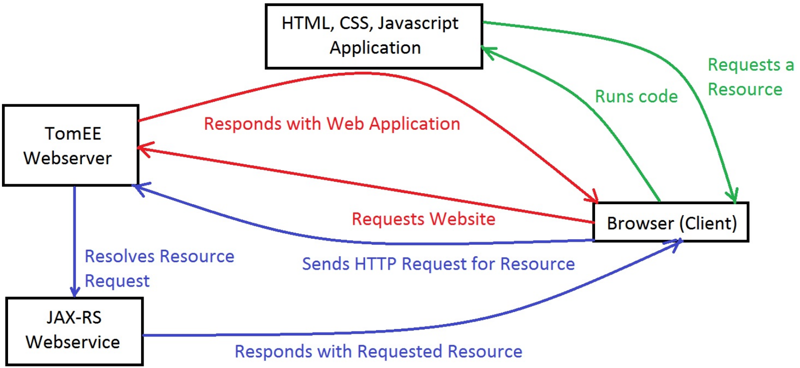

Process
Thread #1
Thread #2
Time

MySQL Database

- Snapshot of the site as of week 1 as HTML files
- Initial work used flat files containing parsed data from site
- Given online goals, database is preferable
- Created a MySQL database containing data collected from the OEIS
- Preferable to get up-to-date data from OEIS, was not practical given time constraints

Full Application

Database-Connected Java RESTful Web Service consumed by Javascript Application using D3



Outline

Antichain Transitions

Viewing Hierarchical Clusterings

- How to consult the edge set efficiently when expanding a node?
- Database simplifies this slightly
- When contracting a cluster, merge edges
- Fast to process, involved to code

Databases and Web Services

- Lack of experience
- Lack of experience
- Transition from local to online
- Using databases from within web applications

Continuing Work

- Refinement of web interface
- Query support
- Labeling and clustering improvements
- Performance, concurrency, and caching

Acknowledgements

- We'd like to thank DIMACS, Dr. Fiorini, and everyone else involved in the DIMACS REU program.

DIMACS

Center for Discrete Mathematics & Theoretical Computer Science
Founded as a National Science Foundation Science and Technology Center

MAA

MATHEMATICAL ASSOCIATION OF AMERICA

References

[1] **The On-Line Encyclopedia of Integer Sequences (OEIS)**, http://oeis.org.

[2] ***Ask Graph View*****: The Design of a Large Scale Graph Visualization System** by J. Abello, F. Van Ham, N. Krishnan, 2006.

[3] ***The State of the Art in Visualizing Dynamic Graphs*** by Beck, Burchm Diseh and Daniel Weiskopt, Appeared in Eurovis 2014.

[4] ***(Semi-)External Algorithms for Graph Partitioning*** and Clustering by Akhremtsev, Sanders, and Schulz. 2014.

Any questions?