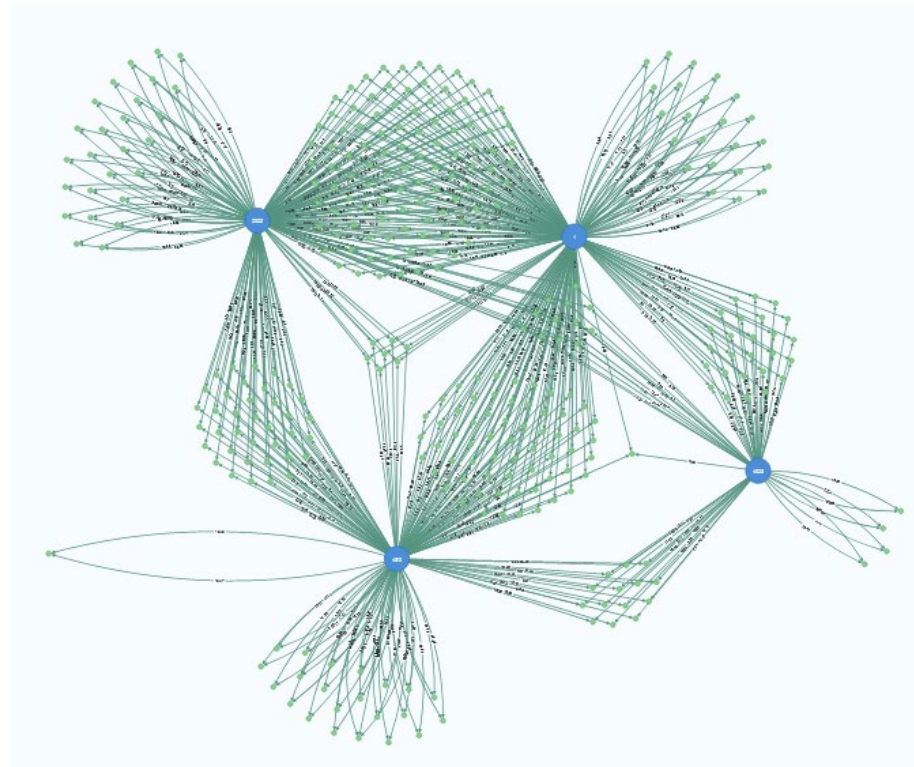


Genealogy Forum of Oregon



**GRAPHS FOR
GENEALOGISTS**

Who Am I, LLC
David A Stumpf, MD, PhD
dave@wai.md

Thinking in graphs will change your world view ...

Migrating to Graph Analytics

... Designed for success!

GRAPHS FOR GENEALOGISTS

- GFG makes the migration to graphs easy
- Familiar genealogy tasks for data preparation
- Step by step guide to set up Neo4j
- GFG Application: download and install
- Load reference and your genealogy data
- Menu driven reports



Videos and Facebook GFG Forum

GRAPHS FOR GENEALOGISTS™

Videos to get you Started

A series of brief videos will guide you through the steps required to implement Graphs for Genealogy

Join us on Facebook in the [GFG Forum](#).



Am I Ready for GFG (in development)

This video helps you assess whether GFG is right for you. You'll see how graph methods are well suited for genealogists and how your analytics will help with your research questions.



Preparing Your Data (12 minutes)

GFG uses data familiar to genealogists. This video helps you get it organized for import into GFG. Once that is done, GFG takes care of the rest.



GFG Software (5 minutes)

This video shows you how to acquire and install a menu driven application for managing projects, loading data and creating reports.



Installing Neo4j and the GFG Plug-In (12 minutes)



Setting up a GFG Project (in development)

This is a Paragraph. Click on



Cluster Matching and GraphXR Visualization (4 minutes)

Set-up video: <https://wai.md/gfg>



An invitation ...

How to Get Involved

An invitation to help build the future of genealogy graph analytics.



Testers

GFG needs testers to shake down new features and provide feedback. You'll get an inside track to the latest features and help influence improvements and new features. To volunteer send us an email.
info@wai.md



Advocates

GFG encourages thought leaders to learn about its capabilities and write reviews for their constituency. To request a demonstration or interview send us an email.
info@wai.md



Developers

GFG is an open source project with modular components. Plug-Ins are coded in Java. If you'd like help design, build or manage new capabilities send us an email.
info@wai.md



Innovators

Have an idea? GFG can help you explore its feasibility and develop a plan. There are many opportunities for entrepreneurs using graphs in genealogy. We'd like to help expand the innovation community.
info@wai.md



Collaborators

Would you like to incorporate graph capabilities into your products? GFG can help. DNA analytics, forensic genealogy, document and image management have significant opportunities.
info@wai.md



Meeting Planners

GFG can provide speakers or workshop leaders who bring graph thinking and methods to your meeting.
info@wai.md

Volunteer: <https://wai.md/gfg>



Evolution of Databases

... from data management to retrieval

- CODASYL Data Model (1959) basic requirements
 - Keys
 - Scans
 - Links
- Relational databases (DB2, Access, SQL Server, Oracle, etc.)
 - Tables
 - Joins (“relationships”)
- NoSQL databases (MongoDB, Cassandra, etc.)
 - key : value pairs
 - Scanning efficiency
 - No tables
- Native graph database (Neo4j, Neptune, etc.)
 - First class relationships (links)
 - Nodes, relationships and properties
 - Graph traversals
 - Little performance change with scaling of data or number of graphs
 - Generally, half the size of relational database

Emphasis on data management

Emphasis on data retrieval



The School of Hard Knocks

... the technology problem

- Genealogists use graphs
 - Family trees
 - FAN groups
 - Haplotrees
 - Migration paths
 - Shared matches
 - Patrilineal trees
 - Clusters
 - Etc., etc.
- Analytics are computationally challenged
 - Relational systems ... not graph friendly
 - Matrix system ... massive size, sparsely populated
 - My Autocluster has 3M cells
 - GFG data would require 62T cells -- untenable
 - Graph methods generally unavailable
 - Managing provenance

All within scope
for GFG



Neo4j Native Graph Database



- Market leader
- Open source
- Community (free) and Enterprise editions
- Local or cloud server
- Multiple drivers: python, .NET, java, etc.
- API or direct server interface
- Starting with version 4
 - Multiple databases
 - Access management at granular level
 - Clusters of virtual machines

Set-up video: <https://wai.md/gfg>



Graph database features

- Node (vertex)
- Relationship (edge)
- Properties: metadata
- Indices
- Data types
 - The usual: string, integer, date, Boolean, etc.
 - Graph specific
 - node, relationship
 - Path
 - List and map
- Aggregation strategy
 - Mathematical: count, sum, ...
 - Traversals: path
 - Collect: list
 - Reduce: concatenated string

Set-up video: <https://wai.md/gfg>



GEDCOM, FTDNA and Other Data

...multiple kits

- Family Finder
 - shared_match relationship
 - match_ancestor relationship (surname list)
- Chr Browser
 - Match_segment relationship
 - File name: kit number
- Curated files
 - Links kit and GEDCOM
 - Links triangulation groups to segments
- Reference Data
 - HapMap, Y-haplotree, SharedCM project, etc.

Set-up video: <https://wai.md/gfg>



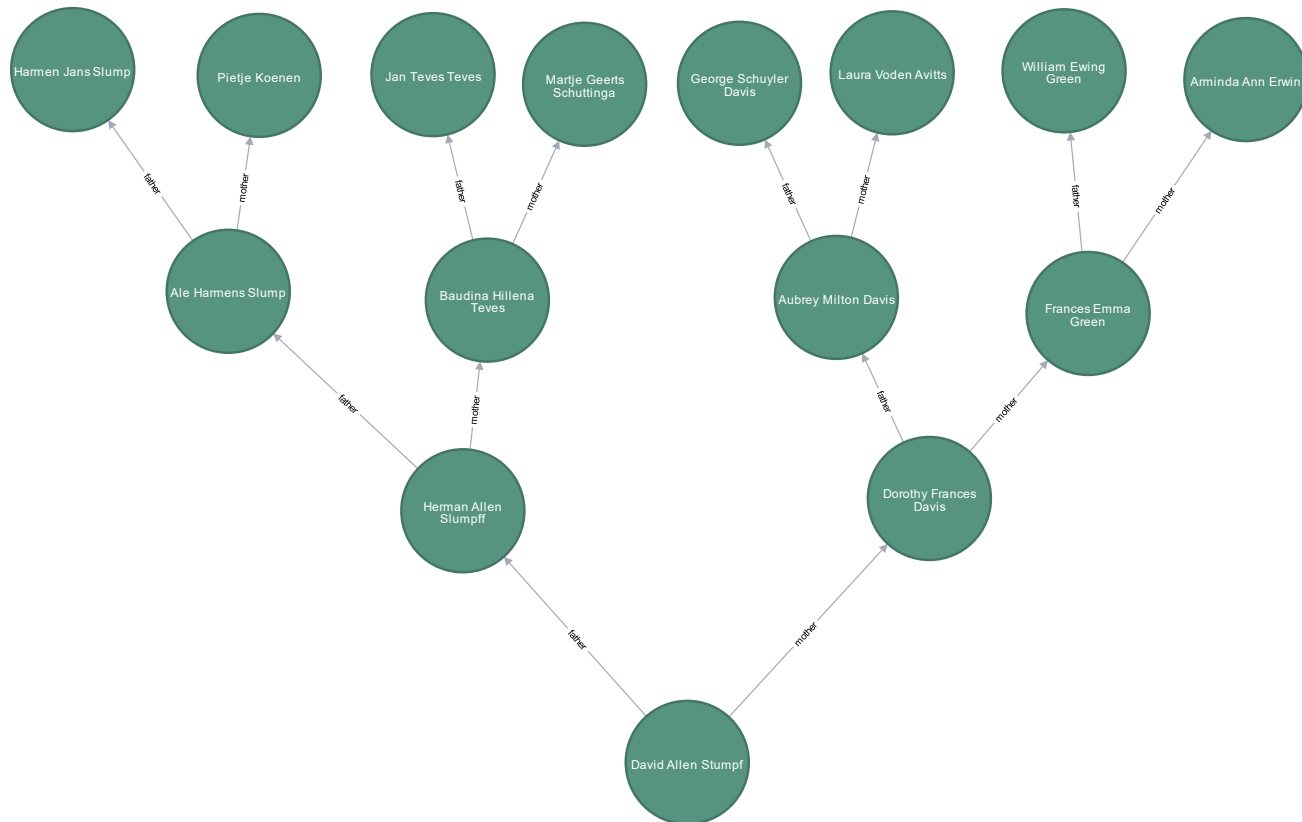
Graph querying

- Graph query language (GQL) standard
- Neo4j Cypher
 - Based on GQL but easier to use
 - Designed to be intuitive
 - Nodes in (): `match (p:Person)`
 - Relationships in []: `match (p1)-[r:father]->(p2)`
 - Properties in {}
`match (p1:Person{RN:27})-[r:father]->(p2)`
 - Query may have a return statement
`match (p:Person{RN:27}) return p.fullname`



Queries are quick and intuitive

... consider how you'd do this in a relational database

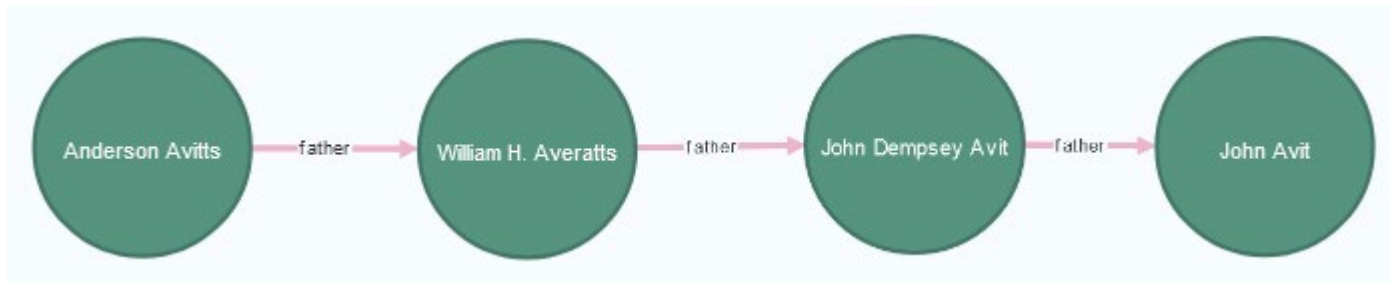


match path=(p:Person{RN:1})-[r:father|mother*0..3]->(a:Person) return path



Graph traversals

- Query finds start node using an index
- Traverses relationships creating paths
- `match path=(p1:Person{RN:27})-[r:father*0..3]->(p2)`
 - Finds person with record number 27
 - Traverses the father relationship through 3 hops
 - Path is returned (not shown in query)



- If you return `p2.fullname` there would be 4 rows.



Collect & Reduce functions

```
match path=(p1:Person{RN:27})-[r:father*0..3]->(p2)
with path, collect(p2.fullname) as c
return c
```

Somewhat analogous
to SQL group by.

Returns 4 rows, one name per row

```
match path=(p1:Person{RN:27})-[r:father*0..3]->(p2)
with collect(p2.fullname) as c
return c
```

Path to List of strings

["Anderson Avitts", "William H. Averatts", "John Dempsey Avit", "John Avit"]

```
match path=(p1:Person{RN:27})-[r:father*0..3]->(p2)
with path, reduce (s="", x in nodes(path) | s + x.fullname + ', ') as c
return c
```

Lists to a string..

Anderson Avitts,
Anderson Avitts, William H. Averatts,
Anderson Avitts, William H. Averatts, John Dempsey Avit,
Anderson Avitts, William H. Averatts, John Dempsey Avit, John Avit,



Ahnentafel number

... built into many reports

```
match path=(p1:Person{RN:27})-[r:father*0..3]->(p2)
with p2, '1' + reduce (s='', x in nodes(path) | s + case when p2.sex='M' then
'1' else '0' end) as c
with p2, '1' + right(c,size(c)-2) as Ahnentafel
return p2.fullname as Name ,
Ahnentafel, gen.rel.ahnentafel(Ahnentafel) as base10
```

Name	Ahnentafel	base10
Anderson Avitts	1	1
William H. Averatts	11	3
John Dempsey Avit	111	7
John Avit	1111	15

Reduce function with a user defined function.



User defined functions (UDF)

- Neo4j supports Plug-ins
 - Java coded functions and procedures
 - Run directly on the server = faster
 - Creates flexibility and enables scalability
- Neo4j plug-ins
 - Awesome procedures for cypher (APOC)
 - Graph data science (GDS)
 - Spatial
- Graphs for Genealogists
 - UDF designed for genealogy data management and analytics
 - Currently 123 functions
 - Open source
 - Enables collaborative development

Open source: <https://github.com/waigitdas/Neo4j-Genealogy-PlugIns>



Motivation for the GFG Plug-In

... success made easy

- Success without in depth graph analytic knowledge
- Enable easy loading of familiar genealogy data
- Robust schema optimized for genealogy analytics
- Application with menu driven reporting
- Focus on discovering actionable insights
- Reports with data + educational information
 - Cypher queries for study and experimentation
 - Explanations using “graph thinking”
- Modular objects supporting re-use
- Scalable, encouraging expert participation in development

Set-up video: <https://wai.md/gfg>



GFG: Making the Complex Simple

```
match (s1:Segment)
with s1 order by s1.chr,s1.strt_pos,s1.end_pos
with collect(distinct s1) as segs1
match (m1:DNA_Match)-[r1:match_segment]-(s2:Segment) where s2 in segs1 and r1.cm>=7 and r1.snp_ct>=500 and
r1.p_anc_rn is not null and r1.m_anc_rn is not null
with r1.p_anc_rn as anc_rn,
case when m1.RN>0 then '*' + m1.fullname + '[' + m1.RN + ']' else m1.fullname end as m2,
case when m1.RN>0 then m1.RN else null end as rns1,
case when r1.m_rn>0 then r1.m_rn else null end as rns2,
case when r1.m_rn>0 then '*' + r1.m + '[' + r1.m_rn + ']' else r1.m end as m3,
max(r1.cm) as max_cm,min(r1.cm) as min_cm,
s2,count(r1) as edgect,sum(case when r1.p=m1.fullname then 1 else 0 end) as unidir_ct_p,
sum(case when r1.m=m1.fullname then 1 else 0 end) as unidir_ct_m
with anc_rn,s2,min_cm,max_cm,
apoc.coll.union(collect(distinct m2), collect(distinct m3)) as matches,
apoc.coll.union(collect(distinct rns1),collect(distinct(rns2))) as rns,
sum(edgect) as edgect,sum(unidir_ct_m) as unidir_ct_m,sum(unidir_ct_p) as unidir_ct_p order by s2.chr,s2.strt_pos,s2.end_pos
with anc_rn,s2,min_cm,max_cm,
apoc.coll.sort(apoc.coll.flatten(matches)) as matches,
apoc.coll.sort(apoc.coll.flatten(rns)) as rns,
edgect,unidir_ct_m,unidir_ct_p
return s2.chr as chr,s2.strt_pos as start_pos,s2.end_pos as end_pos,apoc.math.round(min_cm,1) as cm,edgect,unidir_ct_p,unidir_ct_m,size(rns) as kits,size(matches) as match_ct,anc_rn,rns,matches

matches_by_segments_anc_desc()
```



X-chromosome UDF

...make the complex simple

```
return gen.tree.X_chromosome_inheritance(32)
```

- Prefix “gen” references GFG plug-in
- Function uses a parameter, in this case the record number of a person
- Excel output with five worksheets
 - Ancestors who *may* share an X-chromosome
 - Descendancy tree excluding male to male
 - All X-chromosome relatives
 - Known X-relatives who have DNA test results
 - X-matches in GEDCOM but not in the project (recruitment candidates)

Set-up video: <https://www.wai.md/gfg>



Reduce and ORDPATH in action

Name	gen	sex_path	Ahnentafel	Dewey	ORDPATH
Martha Elizabeth Stinnett [32] (1843-1928)	0	F	1	[32]	1018
Samuel Lewis Stinnett [63] (1822-1864)	1	FM	2	[32,63]	101810127
Alpha Loggins [64] (1823-1906)	1	FF	3	[32,64]	101810128
Mary Elizabeth Nichols [101] (1785-1855)	2	FMF	5	[32,63,101]	1018101271100d
Samuel Loggins [102] (1776-1826)	2	FFM	6	[32,64,102]	1018101281100e
Martha Scott [103] (1795-1880)	2	FFF	7	[32,64,103]	1018101281100f
William Nichols [130] (1754-1850)	3	FMFM	10	[32,63,101,130]	1018101271100d11002a
Nancy Blankenship [131] (1765-1850)	3	FMFF	11	[32,63,101,131]	1018101271100d11002b

- X-chromosome ancestors of Martha Elizabeth Stinnett
- Sex path excludes male to male inheritance
- Dewey is an aggregation of record numbers of persons in the path
- ORDPATH a hexadecimal number created from the Dewey by a UDF
- Notice that sorting on the Dewey will not order the list properly
- Sorting on ORDPATH provides a proper sort
- SQL Server hierarchyid datatype is an ORDPATH string

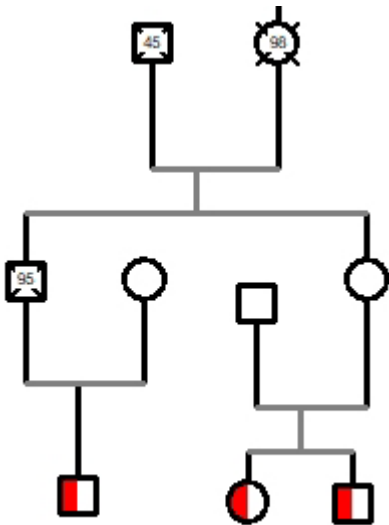
O'Neil, P. *et al.* ORDPATHs: insert-friendly XML node labels. in *Proceedings of the 2004 ACM SIGMOD international conference on Management of data* 903–908 (Association for Computing Machinery, 2004).

<https://www.cs.umb.edu/~poneil/ordpath.pdf>



Common ancestors

```
match (p1:Person{RN:1})-  
[r1:father|mother*0..15]->(mrca:Person)<-[r2:father|mother*0..15]-  
(p2:Person{RN:600})  
return collect(mrca.fullname) as mrca
```



- Pictured pedigree: Query returns grandparents
- Two paths converge on MRCA(s)
- Three variables define a relationship
 - Path length 1
 - Path length 2
 - MRCA count
- Represent as concatenated string
 - Pictured 2:2:2 = 1C
 - 2:2:1 = H1C
 - 4:5:1 = H3C1R
- fam_rel node set also includes Shared Centimorgan Project data

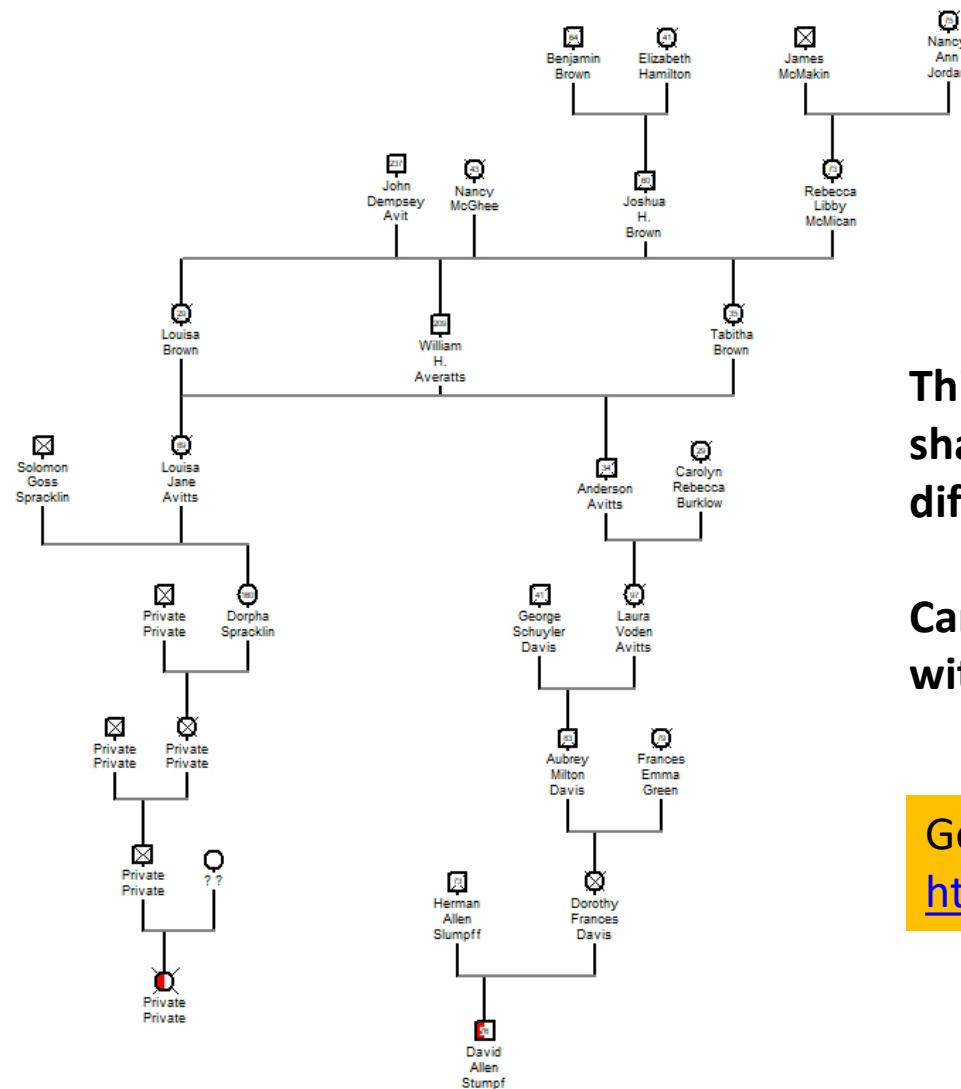
Node Properties

fam_rel	
<id>	52093
HighSharedCM	1225
Indx	2:2:2
LowSharedCM	553
MeanSharedCM	874
nmrca	2
path1	2
path2	2
relationship	1C



MRCAs: it can get complicated

... an easy problem for a graph database



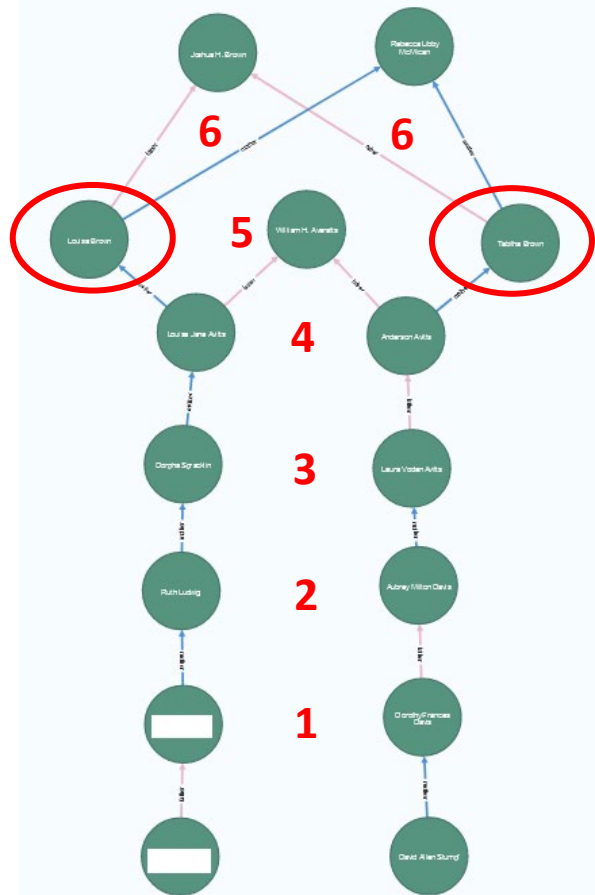
**This relative is my 5C and H4C,
sharing 3 MRCAs in two
different generations.**

**Cannot draw in typical pedigree
without crossing lines.**

Genograms by GenoPro
<https://genopro.com/>



Easier to see in a graph



My 3G grandfather married sisters

His descendant is both my 5C and H4C, sharing 3 MRCAs in two different generations.

5:5:1 = H4C

6:6:2 = 5C

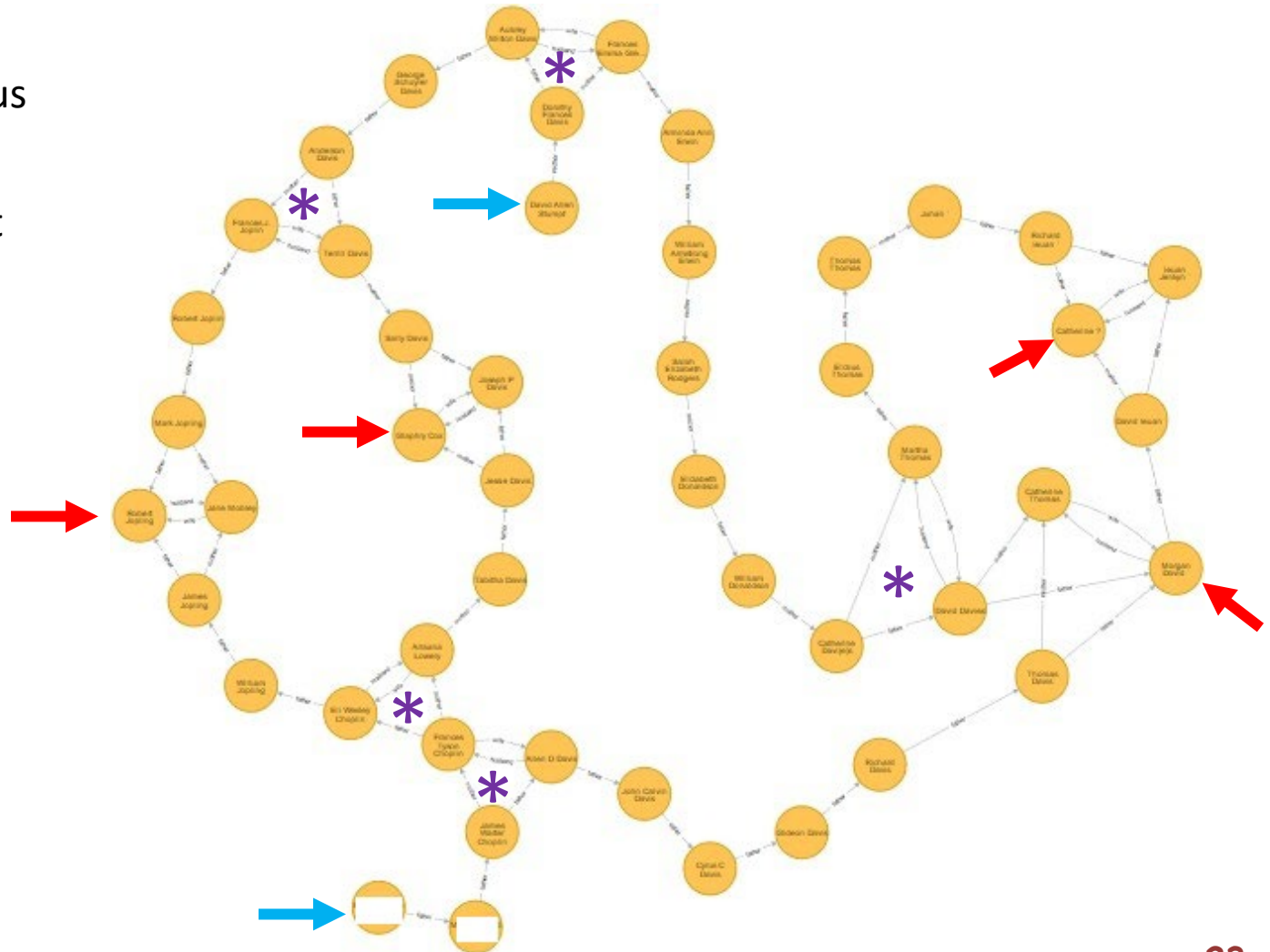
The sisters' children are $\frac{3}{4}$ siblings



Multiple Paths = Ambiguous Match

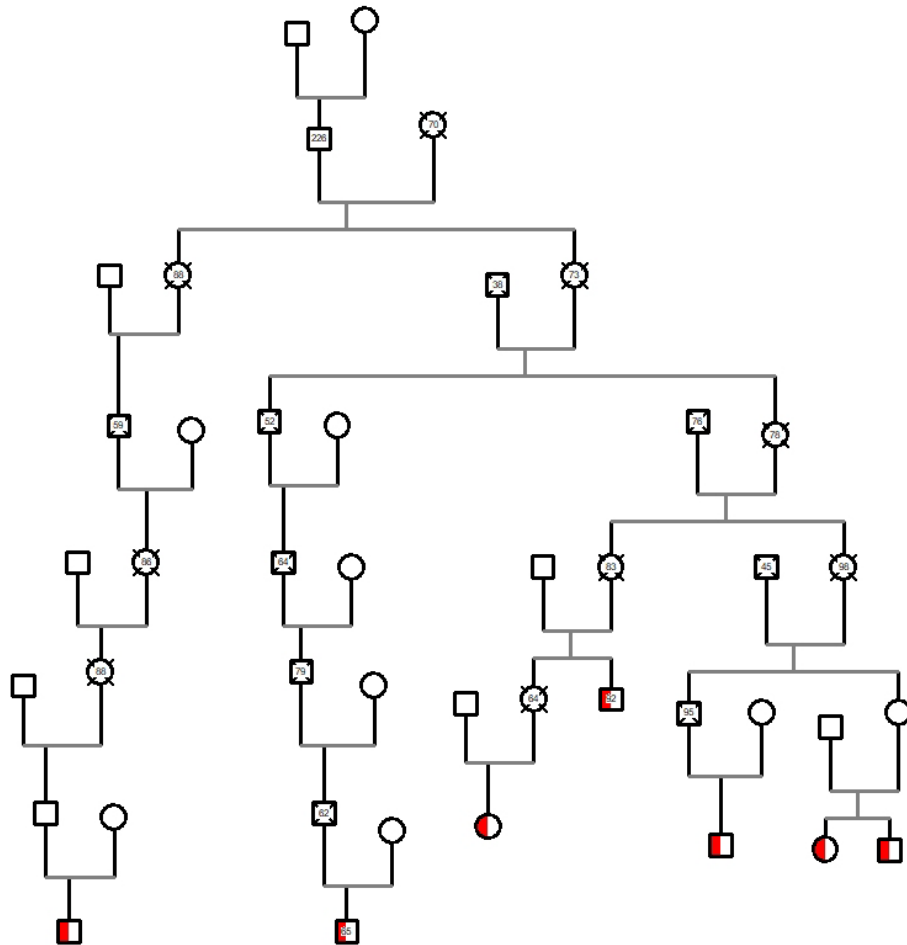
... hard things are easy with a graph database

- ➡ Propositus
- ➡ MRCA
- * Path split



MCRA(s) for a list of relatives

... who is the MRCA of 6 people?



Simple with GFG Function

```
return gen.rel.mrca_from_list('1,52  
42,341,26429, 210')
```

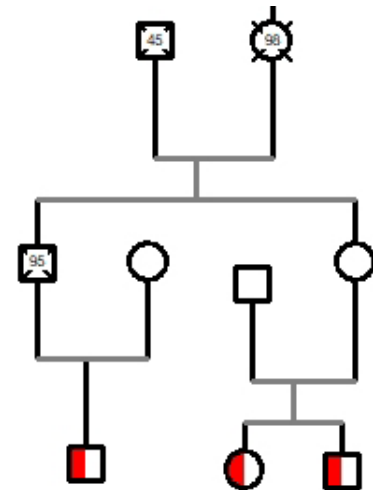
Cypher more complex

```
match (c:Person) where c.RN in  
(['1,5242,341,26429, 210']) With c order by  
c.RN With collect(distinct c.RN) As cc match  
(c2:Person)-[:father|mother*0..20]-  
>(MRCA:Person)<-[:father|mother*0..20]-  
(c3:Person) where c2.RN in cc And c3.RN in cc  
and c2.RN<>c3.RN with MRCA,cc,c2 order by  
c2.RN with MRCA,cc,collect(distinct c2.RN) as  
cc2 with distinct cc,cc2,MRCA.fullname + ' [' +  
MRCA.RN + ']' as CommonAncestor where  
cc2=cc return CommonAncestor
```



Other Pedigree functions

- Patrilineal tree
- Matrilineal tree
- X-inheritance tree
- Double cousins
- Relationships of all in a project
- X-chr genetic distance for all

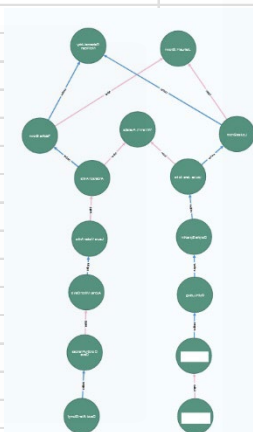


Other Pedigree functions

- Pedigree completeness
- Correlation of relationship (pedigree collapse)
 - Single value
 - Details

propositi	relationship	ancestors	path1	path2	genetic_distance	COR
David Allen Stumpf 1 (1945-) ; anonymized 26429 (-)	H5C	Rebecca Libby McMican 4607 (1790-1863)	6	6	12	.0002441
David Allen Stumpf 1 (1945-) ; anonymized 26429 (-)	H5C	Joshua H. Brown 4441 (1780-1840)	6	6	12	.0002441
David Allen Stumpf 1 (1945-) ; anonymized 26429 (-)	H4C	William H. Averatts 53 (1813-2022)	5	5	10	.0009766
<p>The total COR is 0.00146484375</p> <p>From the shared centimorgan project the expected value and range is unknown cm.</p> <p>The observed shared DNA is 82.9.</p> <p>The predicted DNA is $0.00146484375 \times 6000 = 8.7890625$ cm</p> <p>UDF: return gen.rel.shared_DNA(1,26429)</p> <p>The coefficient of relationship (COR) is a measure of pedigree collapse resulting from ancstors appearing more that one in the family tree.</p> <p>The paths are the generations to the common ancestor for each person in the analysis.</p> <p>references: https://www.yourdnaguide.com/ydgblog/2019/7/26/pedigree-collapse-and-genetic-relationships http://www.genetic-genealogy.co.uk/Toc115570135.html https://isogg.org/wiki/Coefficient_of_relationship database: avitts </p>						

5c
25
0 - 117



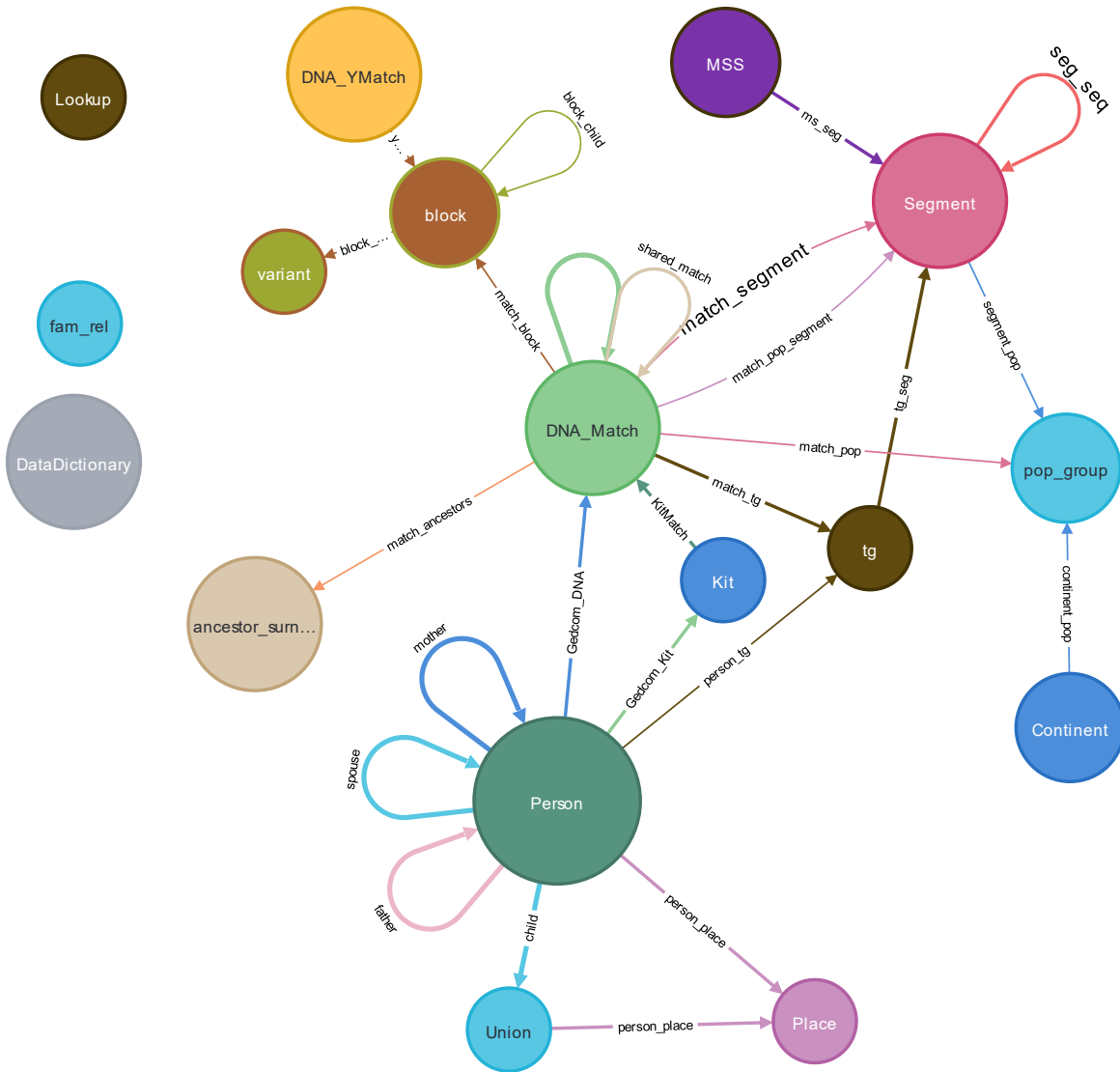
COR, if available, can be a better filter than relationship or centimorgans.

Multiple Graphs

- Up until now ... only the family tree
- Adding each additional graph requires
 - Expert attention to optimizing its graph schema
 - A relationship(s) to other extant graphs
 - Enhancements to optimize query performance
- Properly done ...
 - Meticulously managed provenance
 - Little change in performance (e.g., excellent scalability)
 - Dramatically better analytics (more discoveries)
 - Engages new domain experts (more smart people)
 - Experts have their own toys (app interoperability)



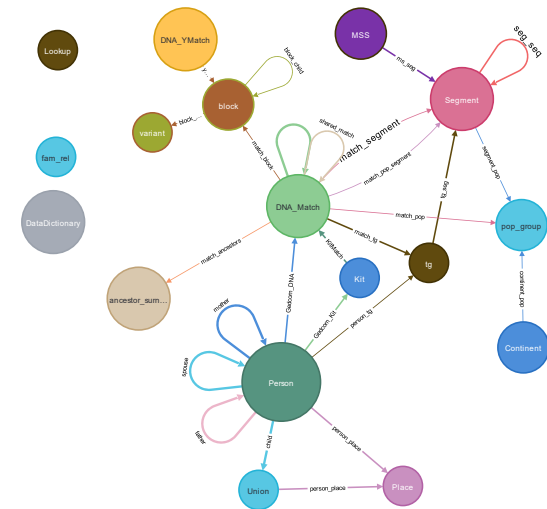
... enhancements using extant data



GFG handles Relatively Big Data

... Stinnett surname project

- Nodes
 - Persons: 34,322 (4,777 Stinnett descendants)
 - Kits: 66
 - DNA_Match: 247,401
 - Segments: 356,721
 - Y-haplotree branches: 54,319
- Relationships
 - GEDCOM_Kit: 42
 - KitMatch: 357,052
 - match_segment: 373,247
 - match_by_segment: 307,364

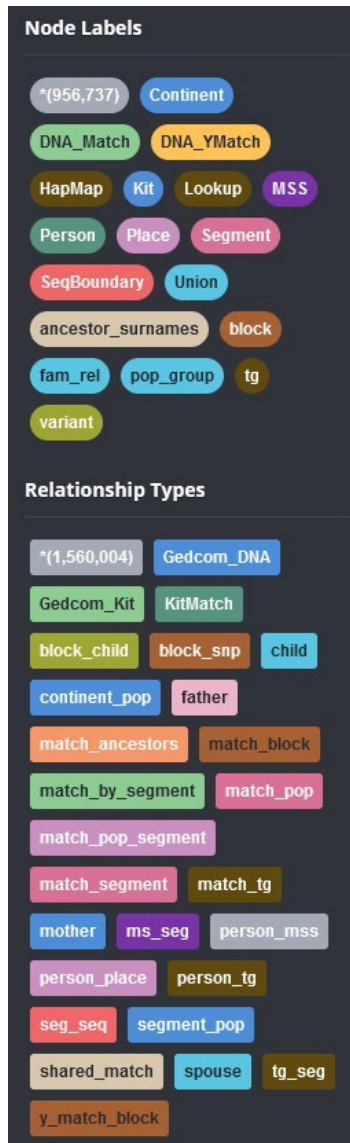


Database size: 825 mB



FTDNA Data Integration

... autosomal DNA



Strategy

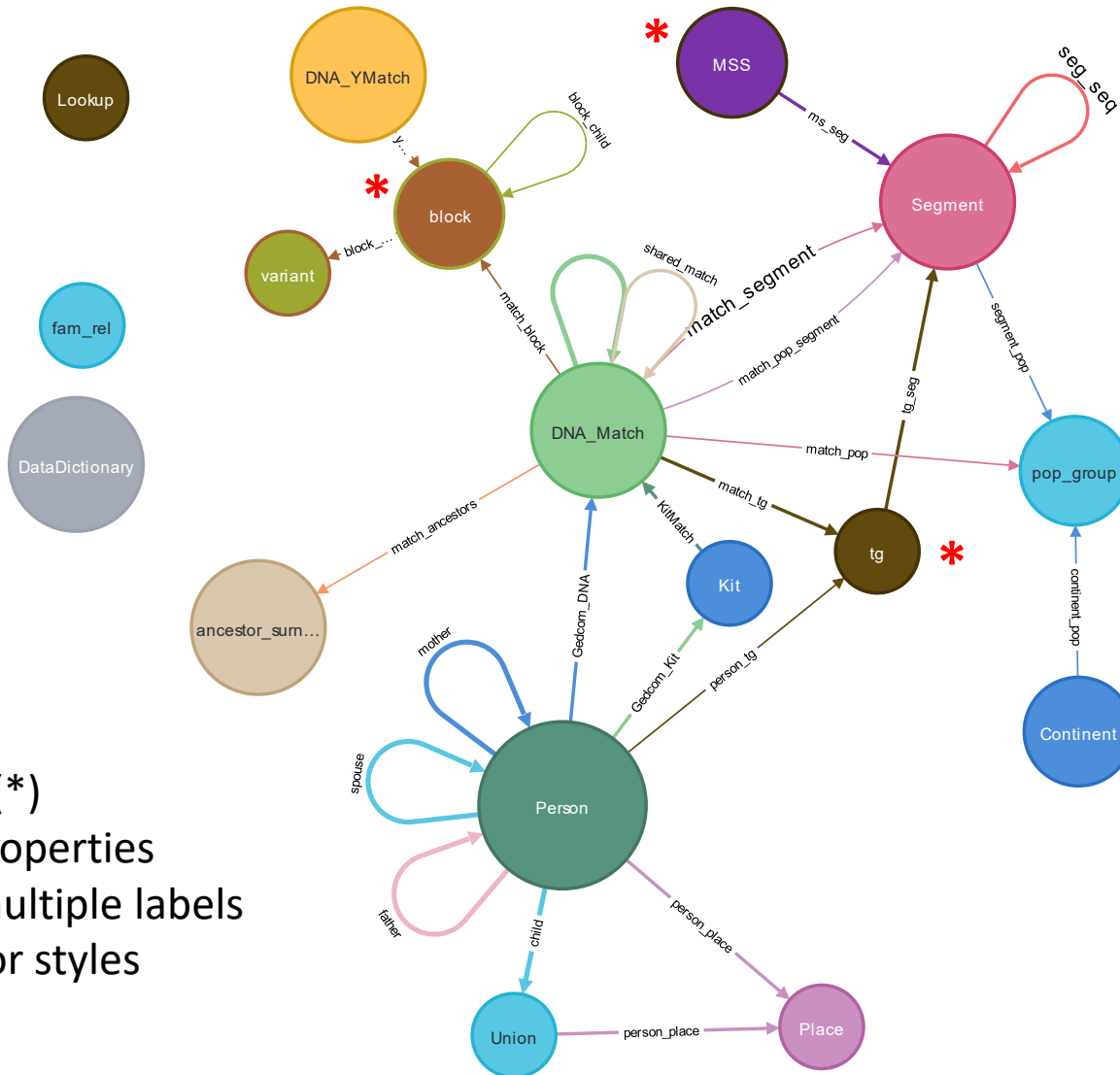
- Preserve original details with sources
- Create Kit, DNA_Match & Segment nodes
- Link these together
- Integrate triangulation group curated data
- Link DNA & family tree graphs together with curated data
- Enhance the graph to optimize analytics

Benefits

- Traversal in 2 or more graphs
- Enhancing graphs by propagating of properties
 - Ancestor RN in nodes (DNA_Match, Segment)
 - Ancestor RN in relationships (match_segment)
- Supernodes (subsume Person, DNA_Match & Segment)
 - Triangulation groups
 - Monophyletic segments
- New relationships from linking analytics
 - Gedcom_DNA, match_by_segment, seg_seq,



Optimized Schema



Super-nodes (*)
Redundant properties
Nodes with multiple labels
*.grass files for styles

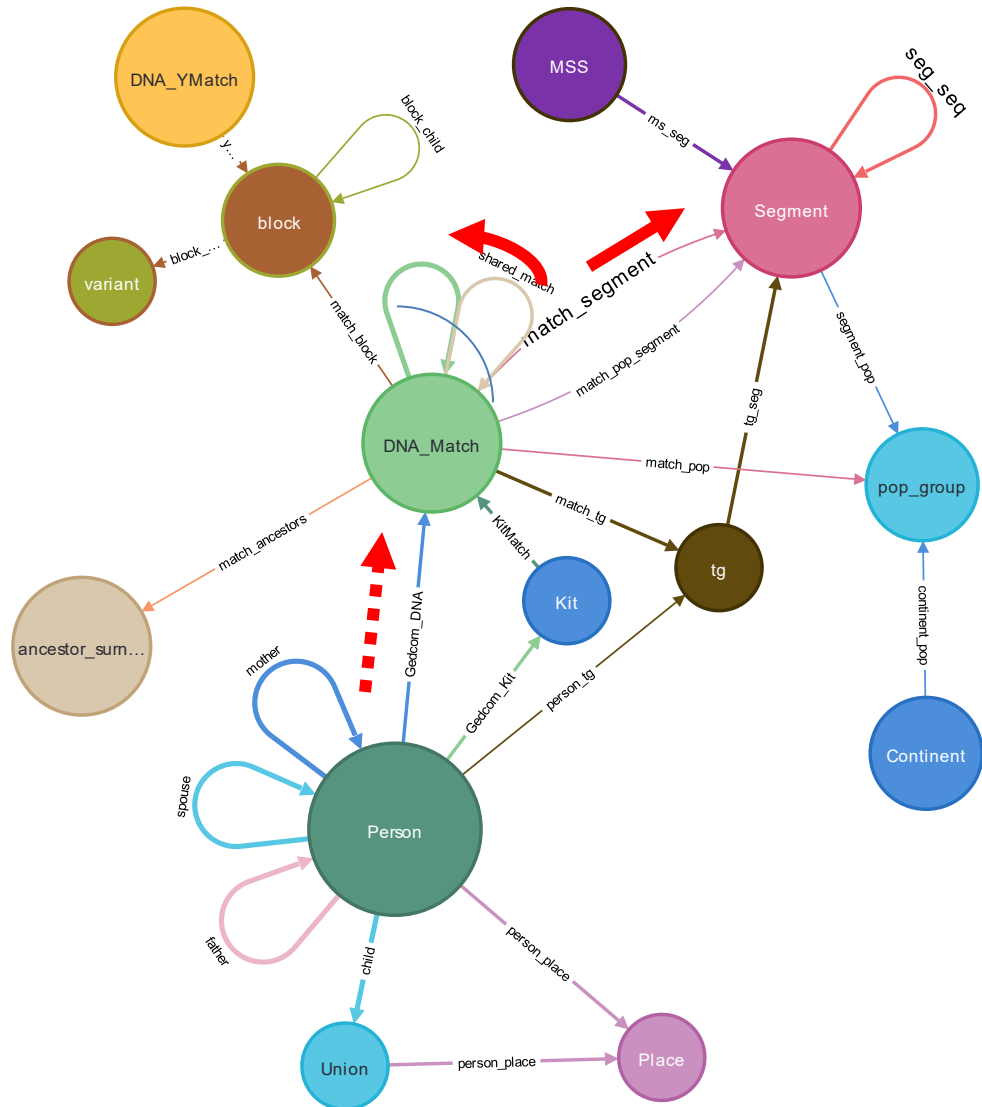


Imagine the Traversals

... graph thinking!



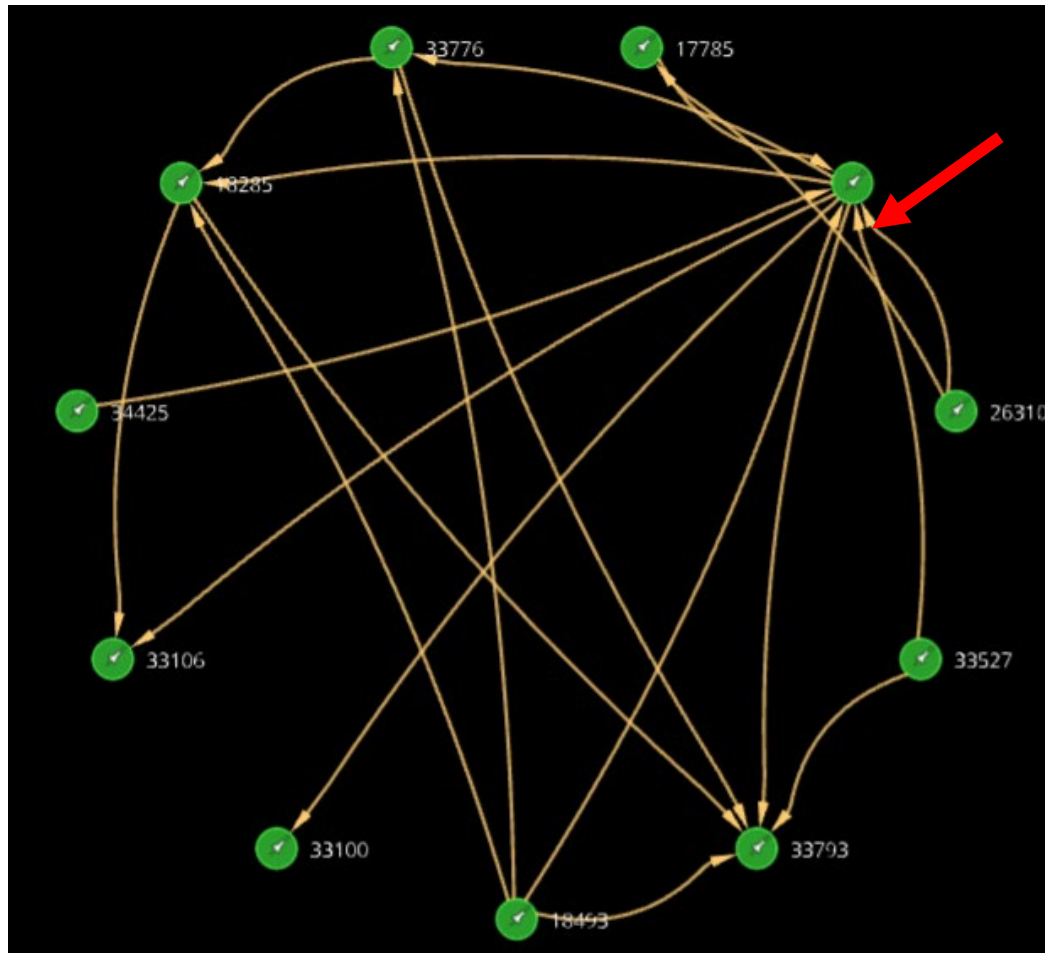
Two useful applications
Cluster matches
Monophyletic segments





Cluster match circle

... DNA_Match nodes linked by match_by_segment



All but one match has a record number.

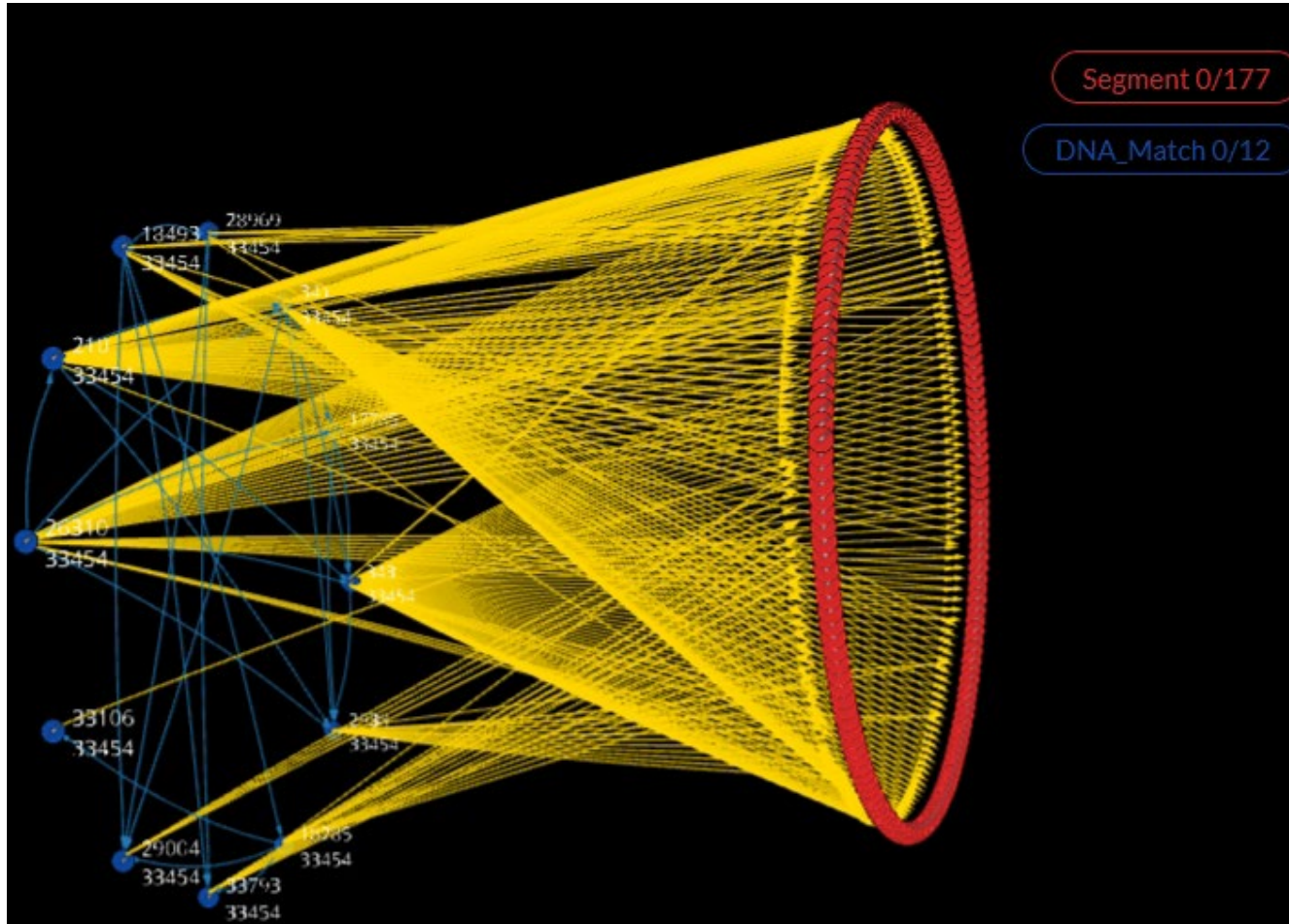
The unknown match (red arrow) has many relationships, providing a motive for additional research.



<https://www.kineviz.com/graphxr>



Ancestor descendant segments



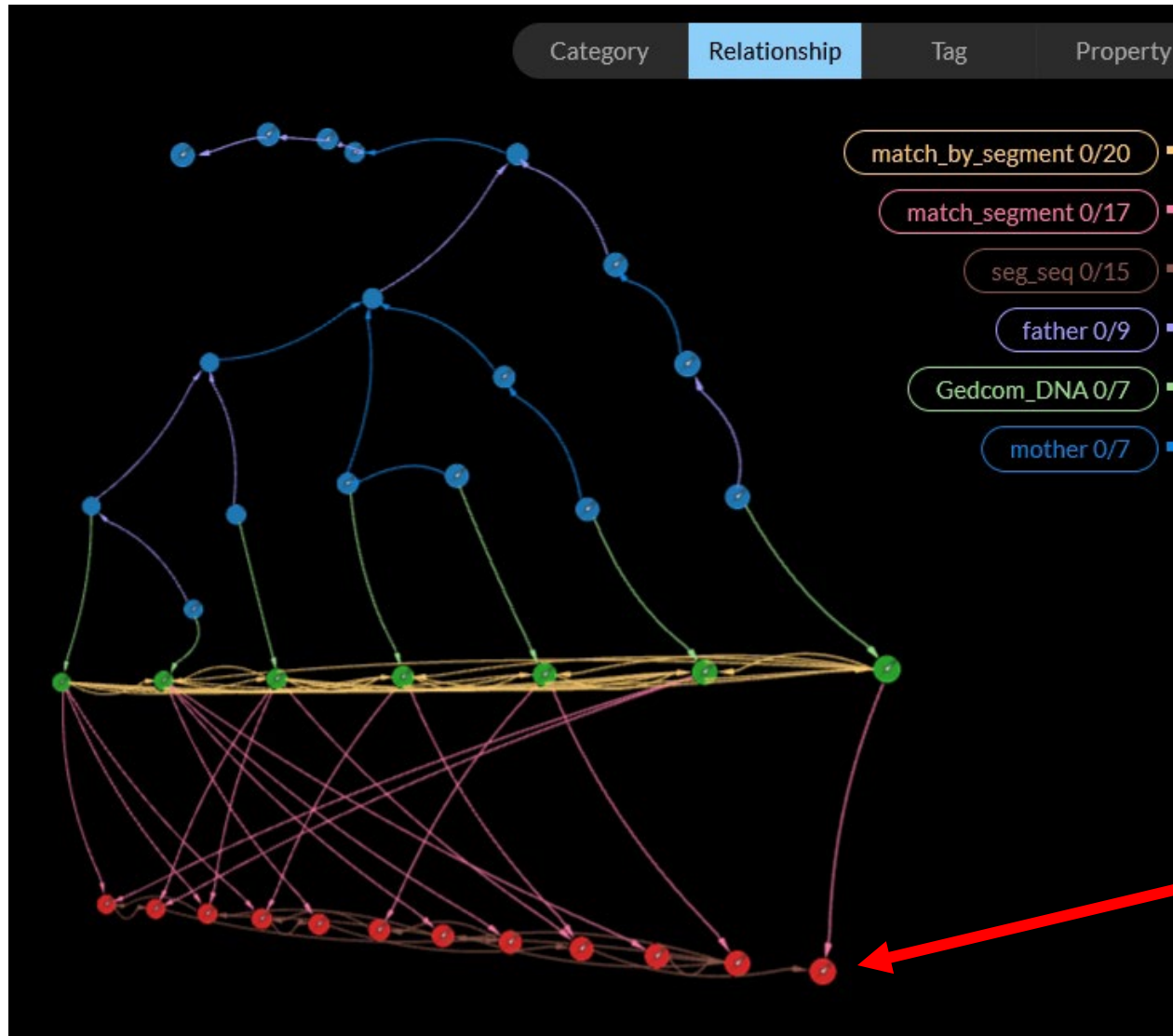
Segments (**red**) and match circle (blue).

<https://www.kineviz.com/graphxr>



at-haplotree

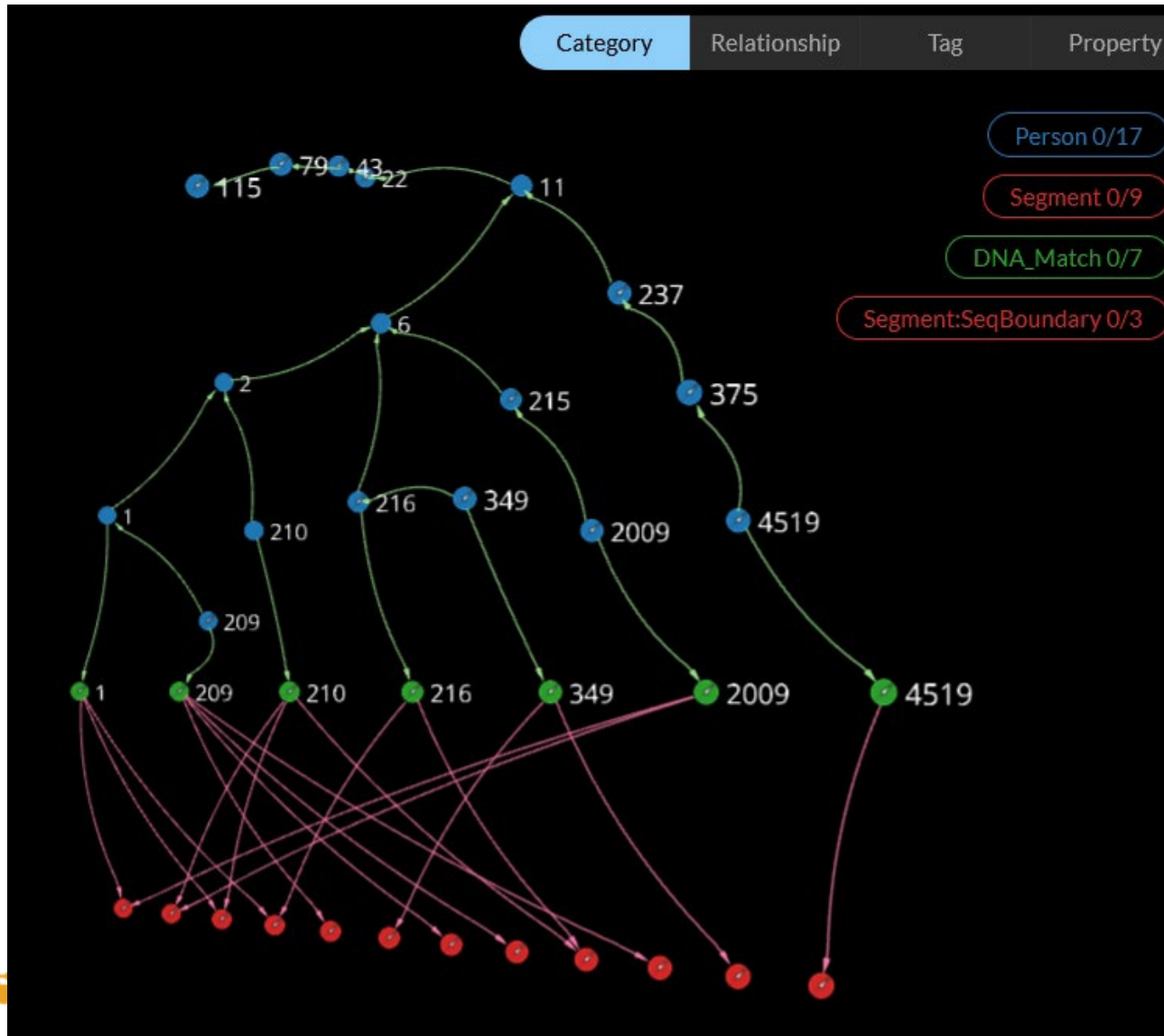
...family tree → matches → segments



Monophyletic segments

at-haplotree

...family tree → matches → segments



Monophyletic segments

... rare and interesting

chr	chr_total_segs	mss_seg_ct	percent
1	31,580	140	0.44
2	30,242	109	0.36
3	20,488	93	0.45
4	22,830	77	0.34
5	19,586	105	0.54
6	21,853	105	0.48
7	18,122	69	0.38
8	15,358	93	0.61
9	17,873	99	0.55
OX	22,391	65	0.29
10	17,331	108	0.62
11	13,969	102	0.73
12	18,089	90	0.5
13	10,286	57	0.55
14	9,791	61	0.62
15	11,888	51	0.43
16	11,617	87	0.75
17	10,649	76	0.71
18	8,510	69	0.81
19	5,572	57	1.02
20	8,602	81	0.94
21	4,891	53	1.08
22	5,259	43	0.82
	356,777	1,890	0.53

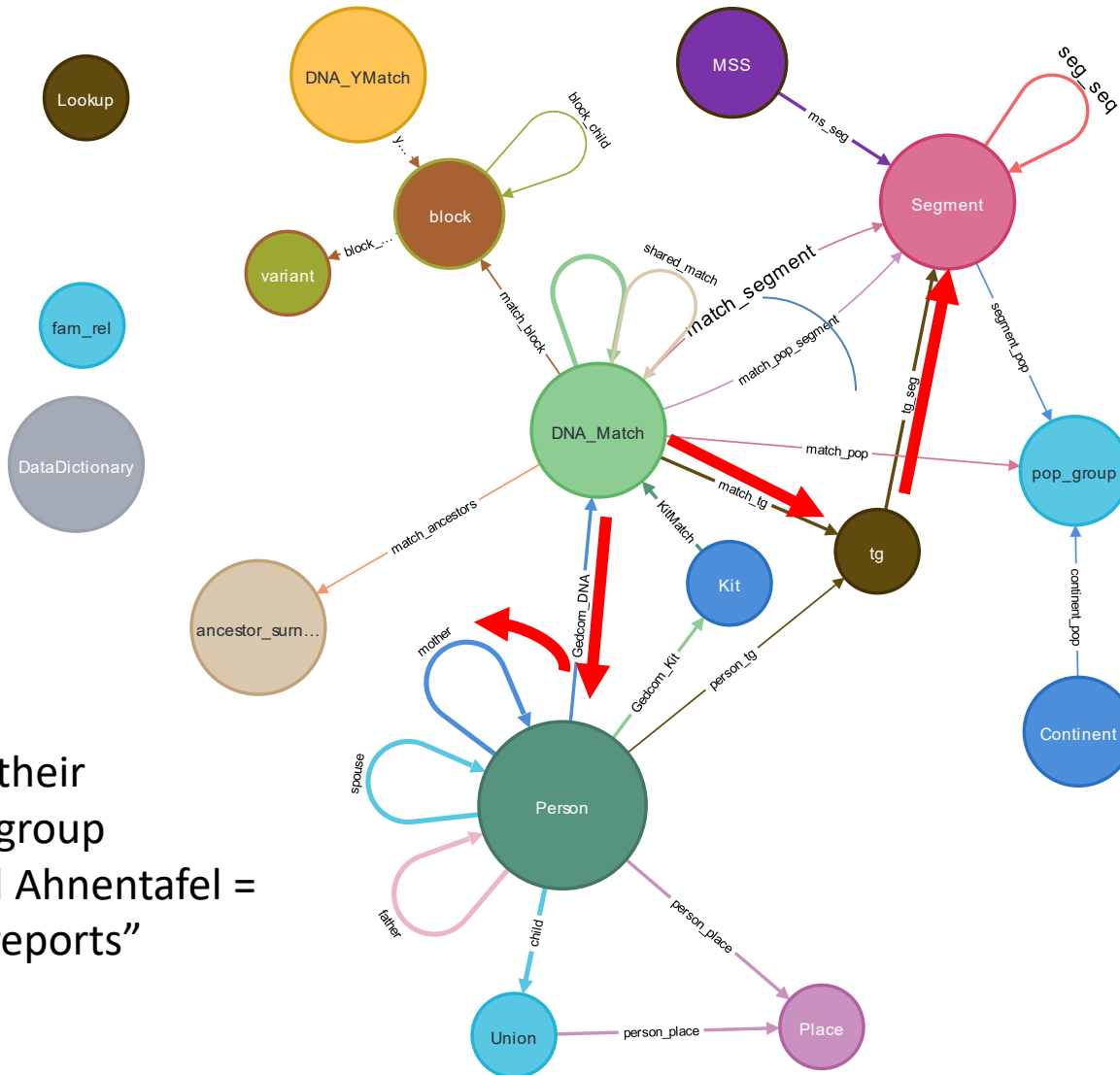
- On average, 1 in 200
- Set of 1,890 Segment nodes = MSS
- Identified by traversals
- Specific family tree paths

mrca	seg	gen	descendant path rns
5443	01:000752566:004913940	4	[33454,5427, 5467 ,5443]
29008	01:000752566:004913940	7	[33454,5427, 5467 ,5443,5436,5609,29008]
29008	01:000752566:004913940	7	[33454,5427, 5467 ,5443,29011,5608,29008]
5467	01:159700039:193826947	3	[33454,5427, 5467]
18256	01:159700039:193826947	6	[33454,5427, 5467 ,5516,5529,18256]

2 segments and their paths
triangulating to 5467



Triangulation Groups

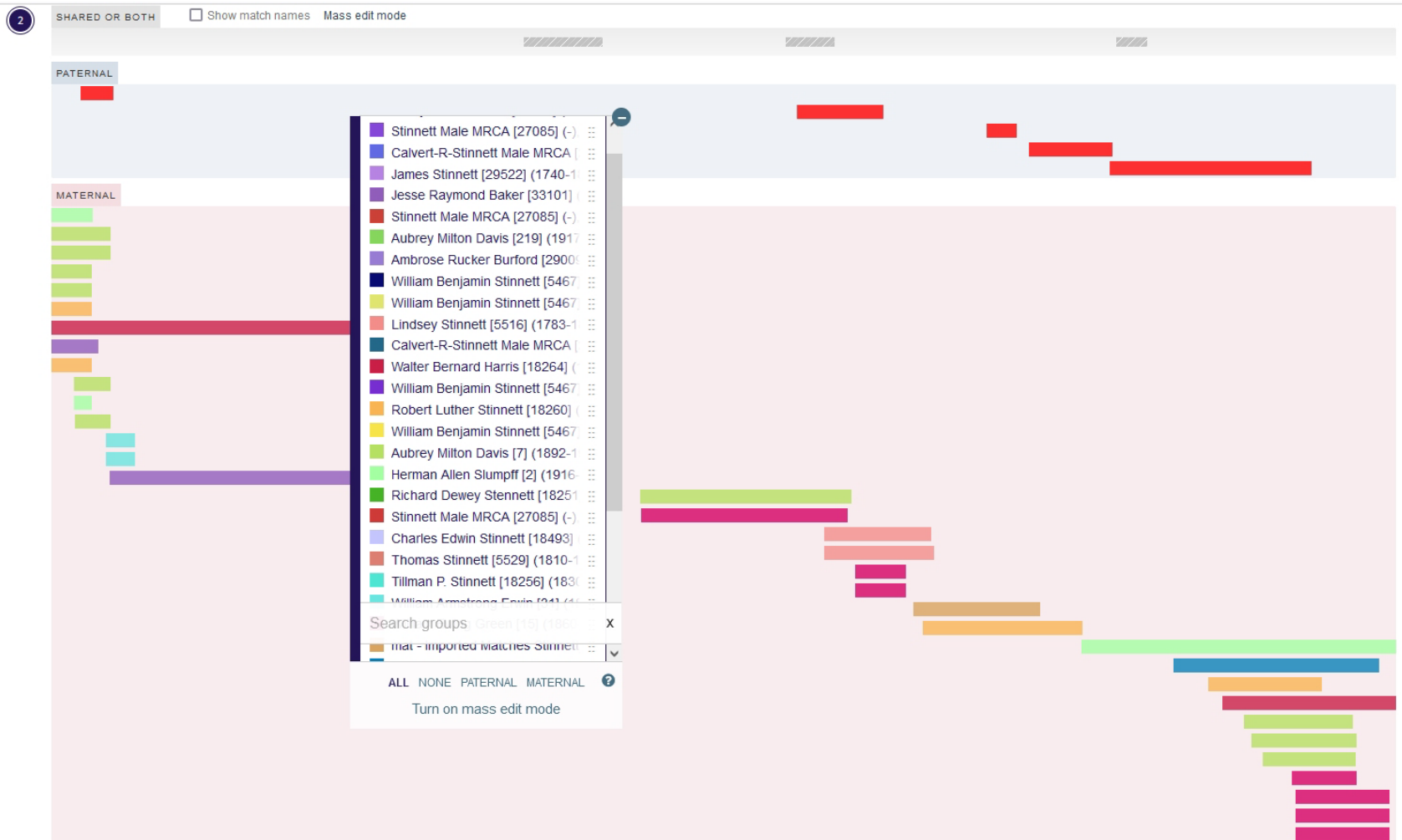


Matches and their
triangulation group
segments and Ahnentafel =
“Jim Bartlett reports”



Triangulation Group: DNA Painter

...some UDFs deliver files for DNA Painter



Brings together family tree, shared matches and segments

at-haplotree

Rendering of the prior graph in Excel

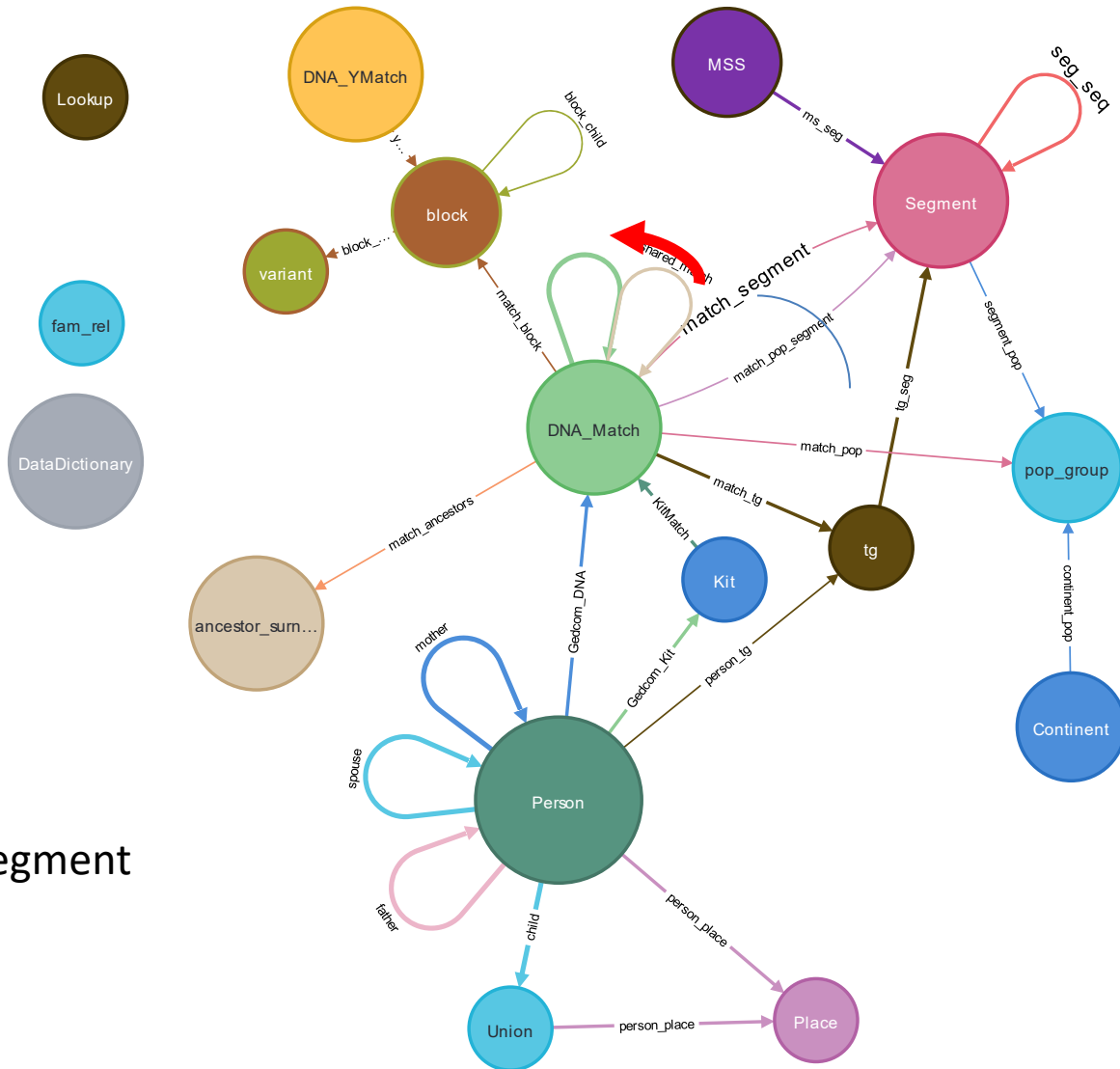
A	B	C	D	E	F	G	H	I
propositus_match	relationship	phased_mrca	chr	start	end	cm	snp	source
	Sibling	[Herman Allen Slumpff [2] (1916-1990)]	20	63,799	59,776,834	102	17637	m
	Uncle-Aunt/Nibling	[Baudina Hillena Teves [6] (1893-1961)]	20	14,840,502	54,178,797	50	9953	p,m
	3C1R	[Jan Cornelis Teves [41] (1796-1833)]	20	51,923,491	58,789,482	18	3052	p,m
	1C	[Baudina Hillena Teves [6] (1893-1961)]	20	52,220,643	62,960,292	29	4955	p,m
	Uncle-Aunt/Nibling	[Baudina Hillena Teves [6] (1893-1961)]	20	14,840,502	54,178,797	50	10651	p,m
	3C1R	[Jan Cornelis Teves [41] (1796-1833)]	20	52,156,847	59,105,451	18	2749	p,m
	1C	[Baudina Hillena Teves [6] (1893-1961)]	20	52,172,404	59,009,023	18	2737	p,m
	Uncle-Aunt/Nibling	[Herman Allen Slumpff [2] (1916-1990)]	20	51,988,626	55,395,329	8	1168	m
	1C1R	[Jan Teves Teves [11] (1854-1913)]	20	55,667,030	61,745,539	18	3023	m
	3C1R	[Jan Cornelis Teves [41] (1796-1833)]	20	36,727,102	42,789,056	9	1303	p,m
	3C	[Teves Jans Teves [21] (1823-1881)]	20	36,945,636	42,712,294	8	1226	p,m
	3C1R	[Jan Cornelis Teves [41] (1796-1833)]	20	29,432,371	43,742,438	15	2957	p,m
	3C1R	[Jan Cornelis Teves [41] (1796-1833)]	20	48,421,680	59,391,936	26	2857	p,m

	A	B	C
1	match	tg_ct	tg
2	David A Stumpf [1]	7	8,17,20,21,22,52,125
3		8	8,17,20,21,22,52,54,125
4		1	20
5		9	8,17,20,21,52,54,102,125,138
6		7	17,20,21,52,102,125,138
7		6	17,20,21,22,102,125
8		9	8,17,20,21,22,54,102,125,138
9		5	8,21,54,102,125
10		7	8,20,21,52,54,102,125
11		1	20
12		1	125

Multiple triangulation
groups create a complex
at-haplotree



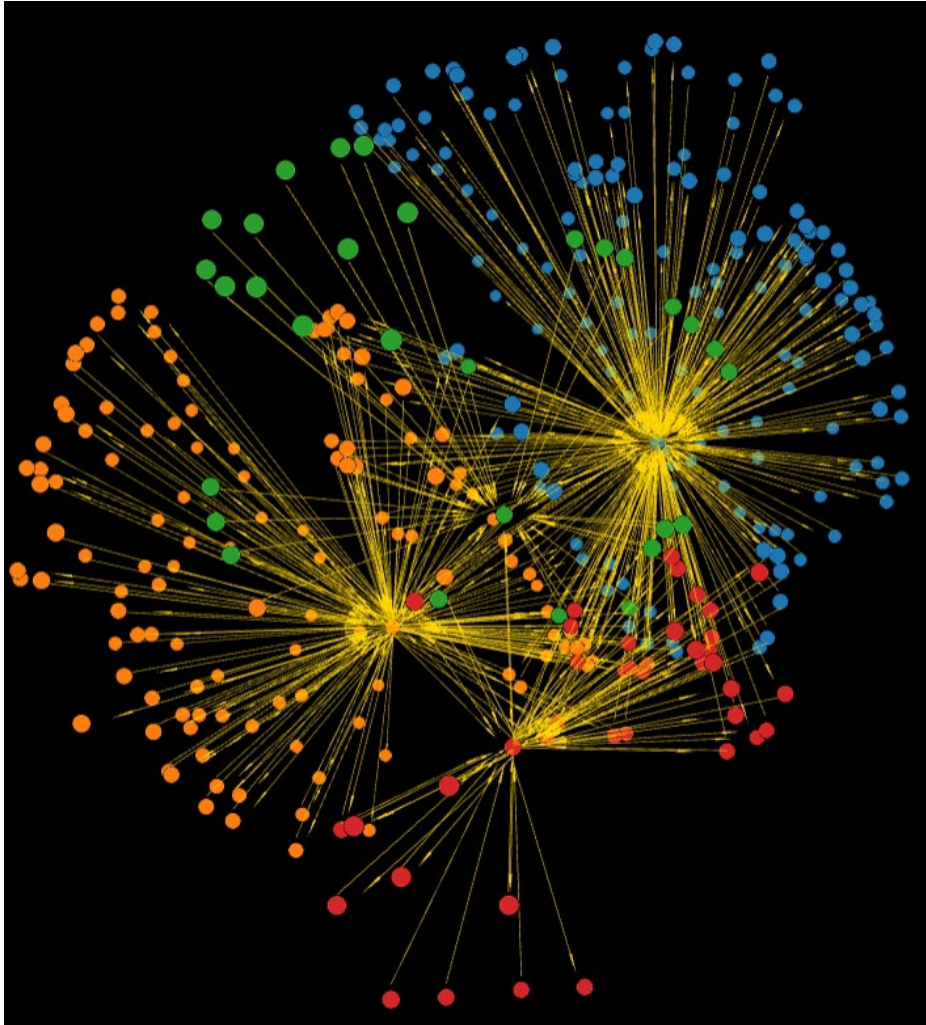
Clusters



Match_by_segment
clusters



Graph Algorithms



GFG uses Neo4j Graph Data Science Plug-In and its community detection algorithms

Louvain

Modularity optimization

Page Propagation

Community members align with family tree branches.

Community segments identified.

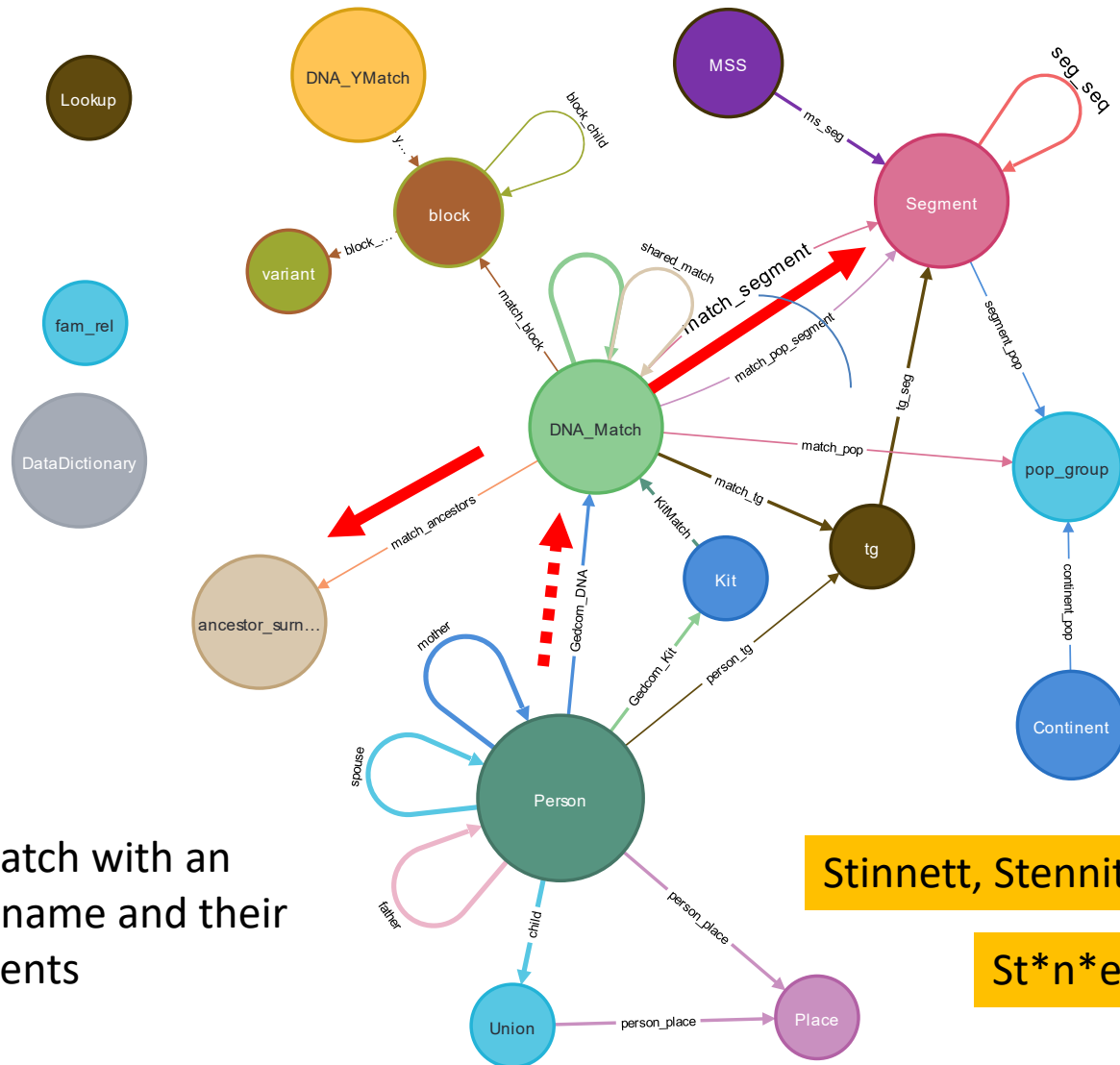
GFG queues up query for ...

DNA Painter

GraphXR (3D renderings)



Matches' Ancestral Surnames



Finding all match with an ancestral surname and their shared segments

Stinnett, Stennit or St*n*et*

St*n*et* or C*I*ert

Ancestral surname reports

... 3 worksheets

- Matches with the surnames
 - Match
 - Source kit
 - Shared cm and segment count
- In-common-with matches
 - Match
 - Cluster of icw matches
 - MRCAs of the cluster of matches, if identifiable
- Shared segments
 - Matches at segments (usually a short list)

CALVERT,KENT
CALVERT,KENT,STINNETT
CALVERT,KENT,STINNETT
CALVERT,KENT
KENT,STINNETT



An invitation ...

How to Get Involved

An invitation to help build the future of genealogy graph analytics.



Testers

GFG needs testers to shake down new features and provide feedback. You'll get an inside track to the latest features and help influence improvements and new features. To volunteer send us an email.
info@wai.md



Advocates

GFG encourages thought leaders to learn about its capabilities and write reviews for their constituency. To request a demonstration or interview send us an email.
info@wai.md



Developers

GFG is an open source project with modular components. Plug-Ins are coded in Java. If you'd like help design, build or manage new capabilities send us an email.
info@wai.md



Innovators

Have an idea? GFG can help you explore its feasibility and develop a plan. There are many opportunities for entrepreneurs using graphs in genealogy. We'd like to help expand the innovation community.
info@wai.md



Collaborators

Would you like to incorporate graph capabilities into your products? GFG can help. DNA analytics, forensic genealogy, document and image management have significant opportunities.
info@wai.md



Meeting Planners

GFG can provide speakers or workshop leaders who bring graph thinking and methods to your meeting.
info@wai.md

Volunteer: <https://wai.md/gfg>

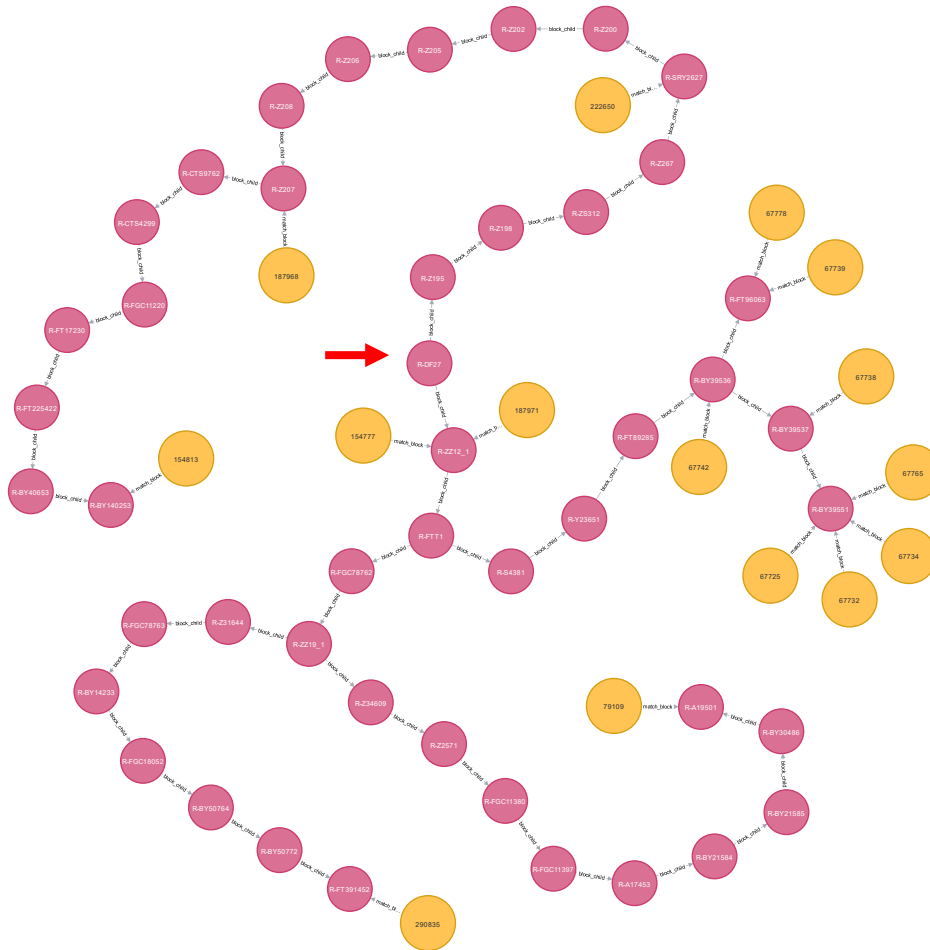


Many Opportunities

- Y-DNA analytics
- mt-DNA analytics
- FAN groups
- Spatial analytics
- Automated DNA Painter interface
- Endogamy
- 3-D renderings
 - Rotating in 2-D plane
 - Virtual reality



Y-Haplotype



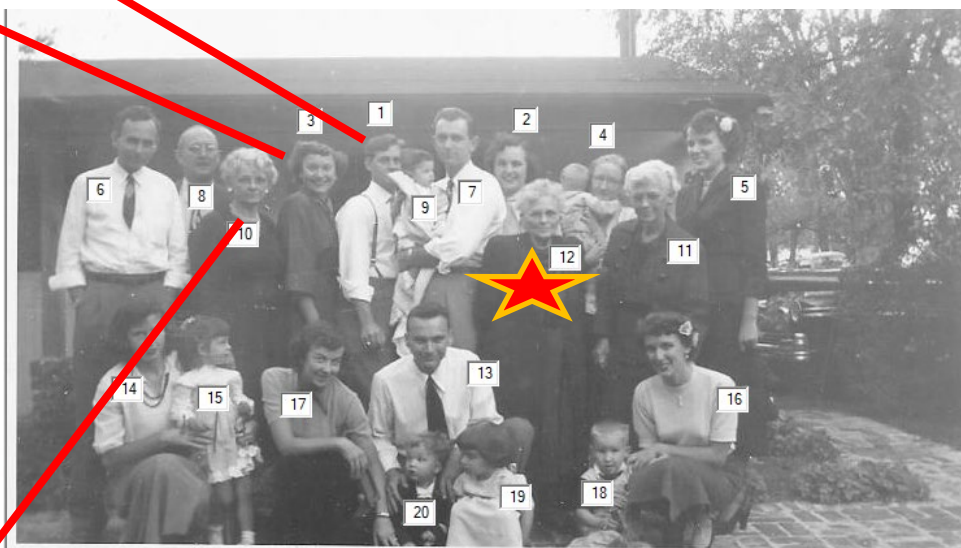
Y-haplotype starting at R-DF29 (arrows) and traversing to snp (red nodes) branch points and kits (orange) with the haplogroups in this branch.

FAN Groups from 1850 Census

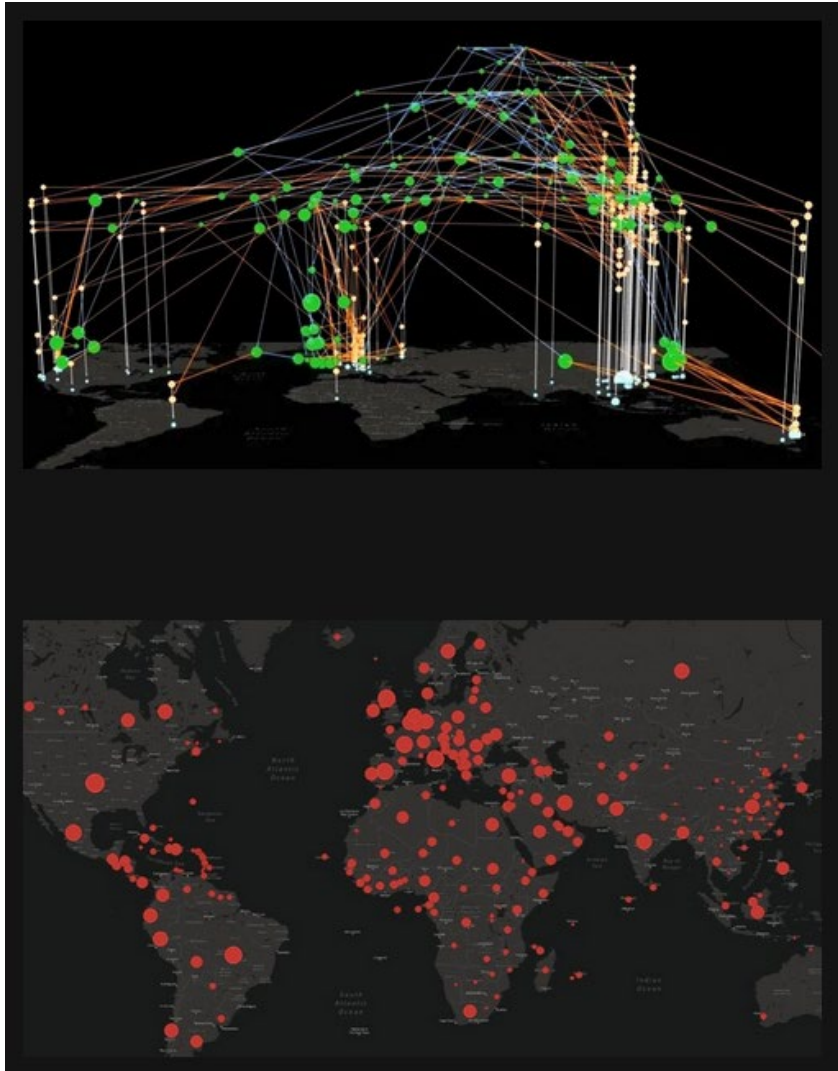


FAN Group from Multiple Sources

[-] Smith, Joseph [26427] (1919-) M :img -
... father: Smith, William Franklin [22779] (1890-1956) M
... wife: Davis, Willie Mae [240] (1889-1967) F
... mother: Avitts, Laura Voden [14] (1868-1966) F
[-] Sorenson, Marie [26428] (1924-) F :img -
... husband: Smith, Joseph [26427] (1919-) M
... father: Smith, William Franklin [22779] (1890-1956) M
... wife: Davis, Willie Mae [240] (1889-1967) F
... mother: Avitts, Laura Voden [14] (1868-1966) F
... Avitts, Anderson [27] (1843-1877) M :Cen - father
... Burklow, Carolyn Rebecca [28] (1844-1873) F :Cen - mother
[-] Avitts, William Henry [270] (1866-1944) M :Cen - brother
... father: Avitts, Anderson [27] (1843-1877) M
... daughter: Avitts, Laura Voden [14] (1868-1966) F
[-] Cain, Margaret Penecia [200] (1840-1907) F :Cen - 36
... husband: Avitts, Anderson [27] (1843-1877) M
... daughter: Avitts, Laura Voden [14] (1868-1966) F
[-] Avitts, Luther Smith [272] (1870-1888) M :Cen - brother
... father: Avitts, Anderson [27] (1843-1877) M
... daughter: Avitts, Laura Voden [14] (1868-1966) F
[-] Avitts, Arthur Anderson [432] (1875-1943) M :Cen - half brother
... father: Avitts, Anderson [27] (1843-1877) M
... daughter: Avitts, Laura Voden [14] (1868-1966) F
... Davis, Aubrey Milton [7] (1892-1976) M :Cen - son
... Davis, George Schuyler [13] (1869-1910) M :Cen - husband
... Davis, Willie Mae [240] (1889-1967) F :Cen - daughter
... Davis, Lessie Wenonah [242] (1890-1988) F :Cen - daughter
... Davis, Rosalie [244] (1896-1972) F :Cen - daughter
... Davis, Patsy Ola [245] (1899-1963) F :Cen - daughter
... Nybye, Julius Oscar [191] (1861-1942) M :Cen - husband



Geography



GraphXR Geospatial & Time Series

Migration paths

Event locations

Surname locations

Point in time family tree

<https://www.kineviz.com/graphxr>



Painted segments

... UDFs generate query for DNA Painter

AAE - MS Female

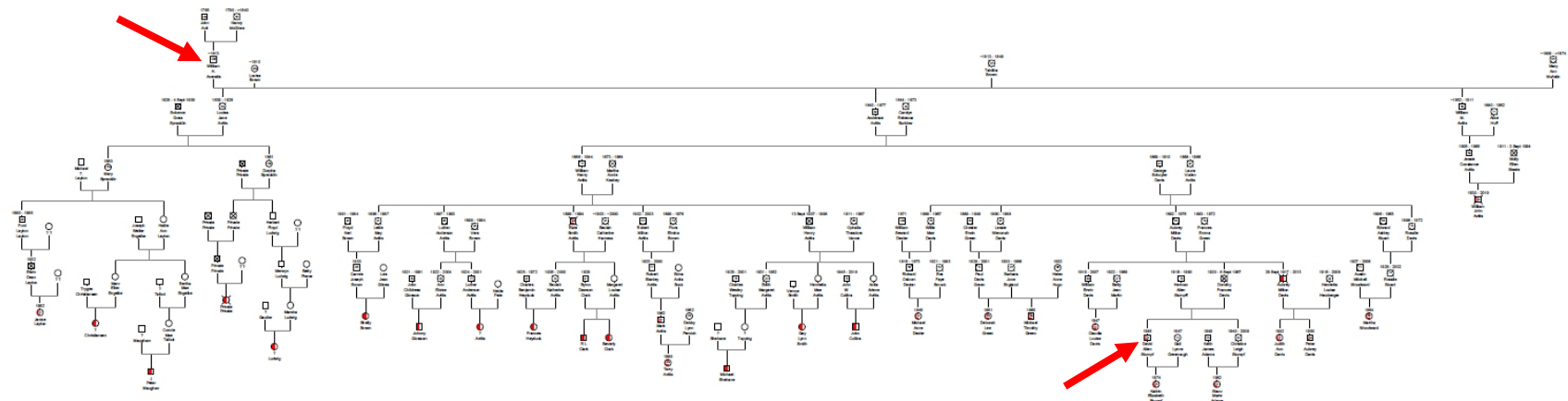
~ 28% / 278 segments painted

DNA PAINTER



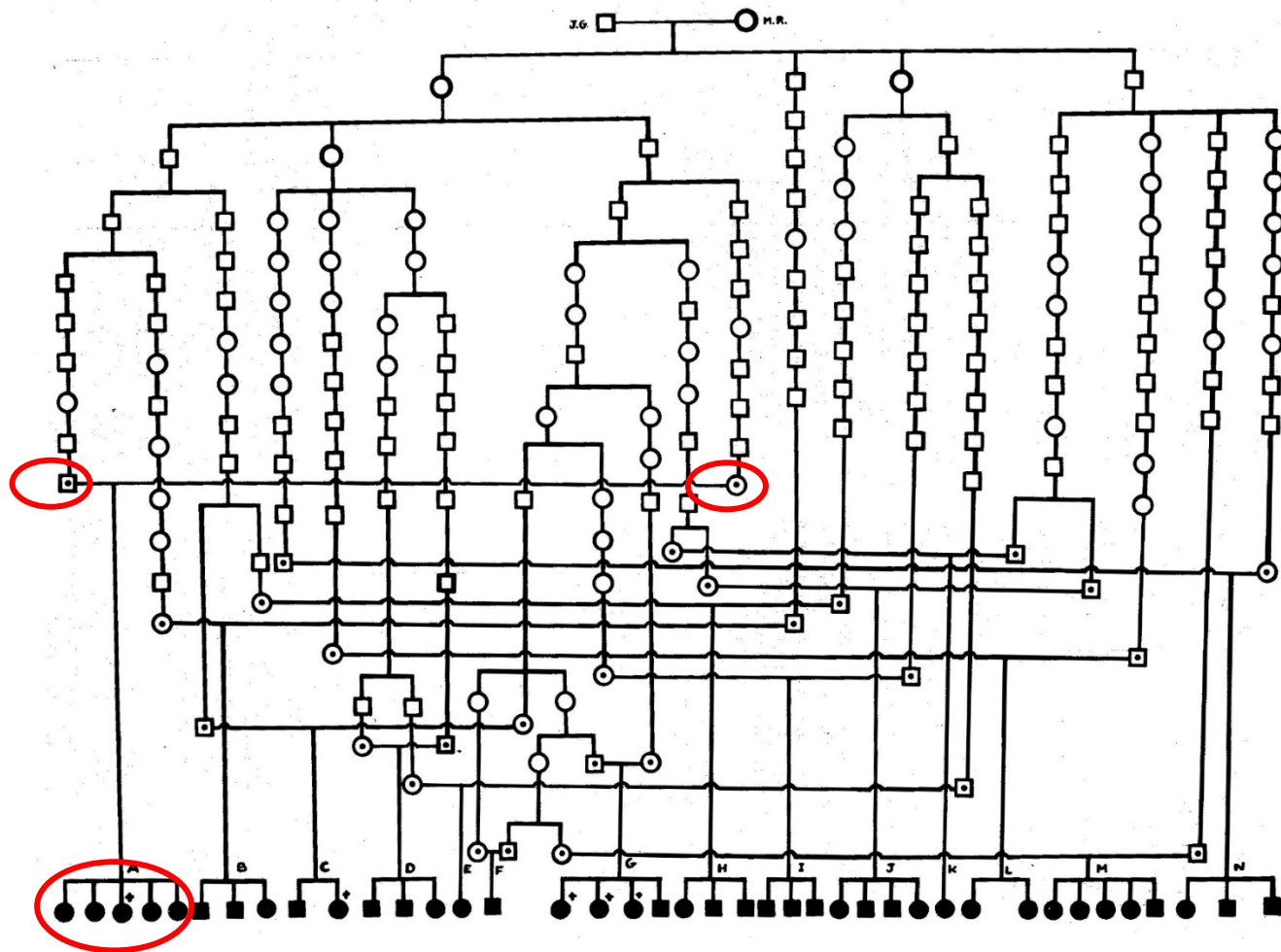
Dendrogram and other charts

MRCA	Match	Name_at_Vendor	RN_Propositus	Shared_CM	Path	Vendor
William H. Averatts [53]	Katrin	Katrin	1	3384	William H. Averatts [53], Anderson Avitts [27], Laura Voden Avitts [14], Aubrey Milton Davis [7], Dorothy Frances Davis	
William H. Averatts [53]	Stacy	Stac	1	2106	William H. Averatts [53], Anderson Avitts [27], Laura Voden Avitts [14], Aubrey Milton Davis [7], Dorothy Frances Davis	23&Me
William H. Averatts [53]	Judith	Judy	1	985	William H. Averatts [53], Anderson Avitts [27], Laura Voden Avitts [14], Aubrey Milton Davis [7], Aubrey Milton Davis [2	FTDNA,ancestry, FTDNA
William H. Averatts [53]	Claudi	Clau	1	970	William H. Averatts [53], Anderson Avitts [27], Laura Voden Avitts [14], Aubrey Milton Davis [7], William Erwin Davis [2	23&Me
William H. Averatts [53]	Peter	P D	1	765	William H. Averatts [53], Anderson Avitts [27], Laura Voden Avitts [14], Aubrey Milton Davis [7], Aubrey Milton Davis [2	FTDNA
William H. Averatts [53]	Micha	Mich	1	518	William H. Averatts [53], Anderson Avitts [27], Laura Voden Avitts [14], Lessie Wenonah Davis [242], Paul Davis Greer	FTDNA
William H. Averatts [53]	Debor	Debr	1	395	William H. Averatts [53], Anderson Avitts [27], Laura Voden Avitts [14], Lessie Wenonah Davis [242], Paul Davis Greer	ancestry.com
William H. Averatts [53]	Martha	MAF	1	249	William H. Averatts [53], Anderson Avitts [27], Laura Voden Avitts [14], Rosalie Davis [244], Rosalie Stuart [390], Marth	FTDNA
William H. Averatts [53]	Elsie	luck	1	161	William H. Averatts [53], Anderson Avitts [27], William Henry Avitts [270], Elsie Smith Avitts [693]	ancestry.com
William H. Averatts [53]	Micha	Mick	1	105	William H. Averatts [53], Anderson Avitts [27], Laura Voden Avitts [14], Willie Mae Davis [240], Richard Osborn Dexter	ancestry
William H. Averatts [53]	Bever	Beve	1	94	William H. Averatts [53], Anderson Avitts [27], William Henry Avitts [270], Elsie Smith Avitts [693], Margaret Louise Avitt	ancestry.com
William H. Averatts [53]	Mark	M. A	1	84	William H. Averatts [53], Anderson Avitts [27], William Henry Avitts [270], Robert Milton Avitts [694], Robert Stanley Avi	FTDNA,FTDNA
William H. Averatts [53]	Terry	T. A.	1	78	William H. Averatts [53], Anderson Avitts [27], William Henry Avitts [270], Robert Milton Avitts [694], Robert Stanley Avi	FTDNA
William H. Averatts [53]	Jean	JS	1	76	William H. Averatts [53], Louisa Jane Avitts [3034], Dorpha Spracklin [17160], Ruth Ludwig [26443], Carl Husemoller [2	ancestry
William H. Averatts [53]	R. L. C	rlcar	1	59	William H. Averatts [53], Anderson Avitts [27], William Henry Avitts [270], Elsie Smith Avitts [693], Margaret Louise Avitt	A
William H. Averatts [53]	Franci	fran	1	42	William H. Averatts [53], Anderson Avitts [27], William Henry Avitts [270], Elsie Smith Avitts [693], Beulah Katherine Avil	ancestry.com
William H. Averatts [53]	Janice	Jane	1	37	William H. Averatts [53], Louisa Jane Avitts [3034], Mary Spracklin [17159], Ford Layton Layton [31465], Elwin Dean L	MH
William H. Averatts [53]	Shelly	Shell	1	37	William H. Averatts [53], Anderson Avitts [27], William Henry Avitts [270], Lettie May Avitts [690], Cammie Joseph Brown	ancestry.com
William H. Averatts [53]	Micha	Mich	1	35	William H. Averatts [53], Anderson Avitts [27], William Henry Avitts [270], William Henry Avitts [697], Edith Margaret Avi	ancestry.com
William H. Averatts [53]	? Avitt	even	1	29	William H. Averatts [53], Anderson Avitts [27], William Henry Avitts [270], Luther Anderson Avitts [692], Luther Anderson	A
William H. Averatts [53]	John C	John	1	20	William H. Averatts [53], Anderson Avitts [27], William Henry Avitts [270], William Henry Avitts [697], Anita Arlene Avitts	A
William H. Averatts [53]	Gay L	Gay	1	18	William H. Averatts [53], Anderson Avitts [27], William Henry Avitts [270], William Henry Avitts [697], Henrietta Mae Avit	ancestry.com
William H. Averatts [53]	Johnn	johnr	1	17	William H. Averatts [53], Anderson Avitts [27], William Henry Avitts [270], Luther Anderson Avitts [692], Ann Eloise Avitt	A
William H. Averatts [53]	William	Willk	5242	0	William H. Averatts [53], William M. Avitts [17091], Jessie Constance Avitts [4706], William John Avitts [5242]	FTDNA
William H. Averatts [53]	? Lud	dgau	1	0	William H. Averatts [53], Louisa Jane Avitts [3034], Dorpha Spracklin [17160], Herbert Floyd Ludwig [30602], Merwyn L	
William H. Averatts [53]	? Chris	wlmk	1	0	William H. Averatts [53], Louisa Jane Avitts [3034], Mary Spracklin [17159], Hattie Ann Layton [29359], Mary Ellen Eng	ancestry.com
William H. Averatts [53]	J Pete	J Pet	1	0	William H. Averatts [53], Louisa Jane Avitts [3034], Mary Spracklin [17159], Hattie Ann Layton [29359], Bertha Mae En	
William H. Averatts [53]	Aubrey	A Da	1	0	William H. Averatts [53], Anderson Avitts [27], Laura Voden Avitts [14], Aubrey Milton Davis [7], Aubrey Milton Davis [2	23&Me
William H. Averatts [53]	David	Davi	1	0	William H. Averatts [53], Anderson Avitts [27], Laura Voden Avitts [14], Aubrey Milton Davis [7], Dorothy Frances Davis [3], David Allen Stumpf [1]	ancestry.com, FTDNA, 0, MH, MyHeritage, 4



The Origins of Friedreich's Ataxia in Quebec

... path-specific monophyletic segments?



A Barbeau, et al.: Can J Neurol Sci 1984; Nov; 11: 4 Suppl: 506-9
<https://pubmed.ncbi.nlm.nih.gov/6391645/>



Summary

- Graphs are everywhere
- GFG graphs are optimized for genealogists
- Multiple capabilities in one package
- Well suited for personal and client research
- We hope you enjoy and use it!
- We welcome your thoughts & questions?
 - Feedback
 - Suggested new capabilities
 - Participation



David A Stumpf, MD, PhD

dave@wai.md

<https://wai.md/gfg>

<https://facebook.com/groups/gfgforum>

