Overview in General

This file contains a general overview of the data in the graph including node labels and relationships types.

References

- jqassistant
- Neo4j Python Driver

Node Labels

Table 1a - Highest node count by label combination

Lists the 30 label combinations with the highest number of nodes. The labels with the lowest node count are listed in table 1b. The total list would sum up to the total number of labels (100%).

The whole table can be found in the CSV report `Node_label_combination_count`.

	nodeLabels	nodesWithThatLabels	nodesWithThatLabelsPercent
0	[Git, Change]	248143	73.639275
1	[Java, ByteCode, Member, Method]	13201	3.917548
2	[Java, ByteCode, Parameter]	13016	3.862647
3	[Git, Commit]	12194	3.618709
4	[File, Git]	10694	3.173567
5	[Java, ByteCode, ParameterizedType, Bound]	7172	2.128373
6	[Java, ByteCode, Bound]	7113	2.110864
7	[Java, ByteCode, Member, Field]	3503	1.039555
8	[Java, ByteCode, Bound, WildcardType]	2908	0.862982
9	[Java, Value, ByteCode, Annotation]	2822	0.837461
10	[Xml, Element]	2144	0.636257
11	[Java, ByteCode, Member, Constructor, Method]	2043	0.606284
12	[Xml, Text]	1436	0.426149
13	[Java, ByteCode, Bound, TypeVariable]	1096	0.325251
14	[Java, ByteCode, Member, Method, Lambda]	938	0.278362
15	[Type, File, Java, ByteCode, ResolvedDuplicate...	900	0.267085
16	[Type, File, Java, ByteCode, Class]	861	0.255512
17	[Type, File, Java, ByteCode, JavaType]	748	0.221978
18	[Java, Value, ByteCode, Primitive]	650	0.192895
19	[Java, ByteCode, Member, Method, GenericDeclar...	569	0.168857
20	[Json, Key]	560	0.166186
21	[Value, Json, Scalar]	544	0.161438
22	[Type, File, Java, ByteCode, ExternalType]	396	0.117518
23	[Author, Git, Person]	295	0.087545
24	[Value, Array]	267	0.079235
25	[Committer, Git, Person]	250	0.074190
26	[Type, File, Java, ByteCode, Class, GenericDec...	236	0.070036
27	[Value, Property]	202	0.059946
28	[Type, File, Java, ByteCode, Interface]	187	0.055494
29	[Java, Value, ByteCode, Class]	172	0.051043

Chart 1a - Highest node count by label combination

Values under 0.5% will be grouped into "others" to get a cleaner plot. The group "others" is then broken down in Chart 1b.

```
<Figure size 640x480 with 0 Axes>
```

Nodes per label combination (more than 0.5% overall)

73.64% (248143)

0.61% (2043)
0.64% (2144)
0.84% (2822)
0.86% (2908)
1.04% (3503)
2.11% (7113)
2.13% (7172)
3.17% (10694)
3.57% (12018)
3.62% (12194)
3.86% (13016)
3.92% (13201)

Legend:
- ['Git', 'Change']
- ['Java', 'ByteCode', 'Member', 'Method']
- ['Java', 'ByteCode', 'Parameter']
- ['Git', 'Commit']
- others
- ['File', 'Git']
- ['Java', 'ByteCode', 'ParameterizedType', 'Bound']
- ['Java', 'ByteCode', 'Bound']
- ['Java', 'ByteCode', 'Member', 'Field']
- ['Java', 'ByteCode', 'Bound', 'WildcardType']
- ['Java', 'Value', 'ByteCode', 'Annotation']
- ['Xml', 'Element']
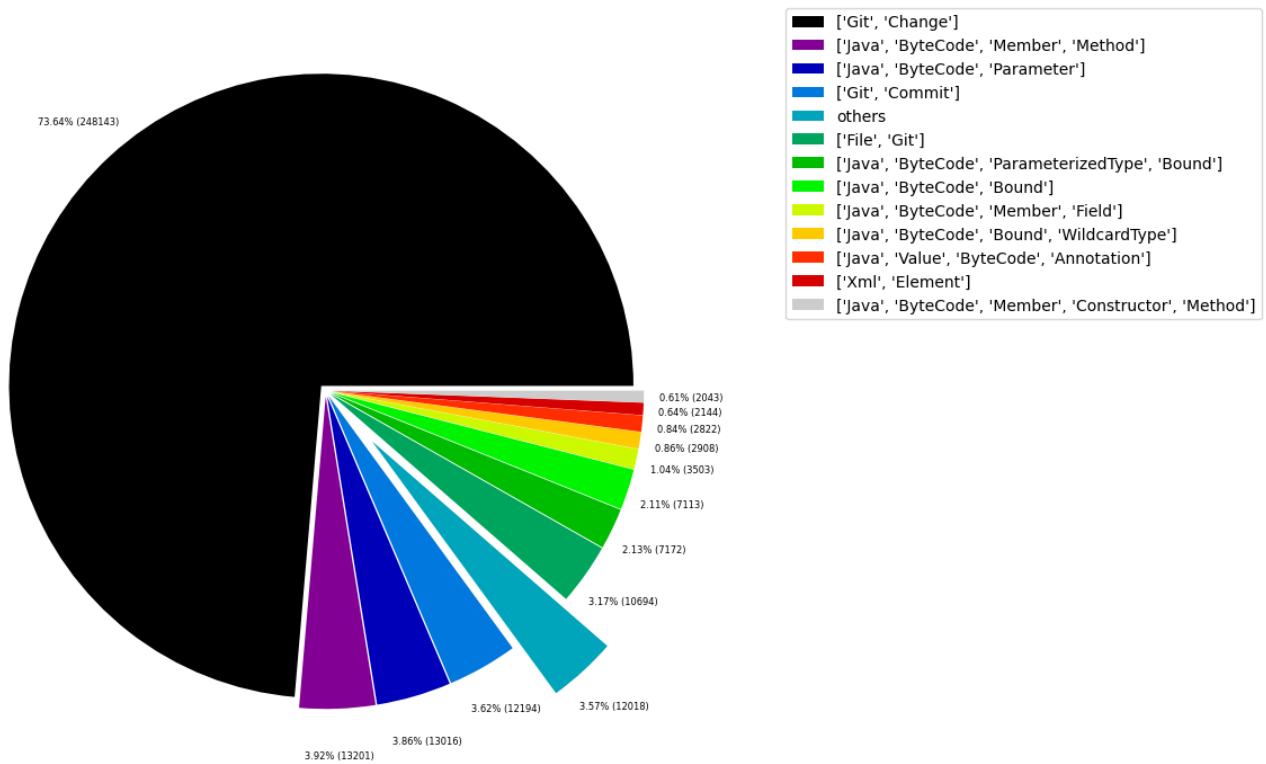- ['Java', 'ByteCode', 'Member', 'Constructor', 'Method']

Table 1b - Lowest node count by label combination

Lists the 30 label combinations with the lowest number of nodes until they reach 0.5% of the total node count, which are shown above.

	nodeLabels	nodesWithThatLabels	nodesWithThatLabelsPercent
0	[Analyze, Task, jQAssistant]	1	0.000297
1	[Repository, File, Git]	1	0.000297
2	[File, Json]	2	0.000594
3	[File]	3	0.000890
4	[Java, ByteCode, Member, Constructor, Method, ...	4	0.001187
5	[Maven, Exclusion]	5	0.001484
6	[Value, Json, Array]	6	0.001781
7	[File, Maven, Xml, Pom, Document]	9	0.002671
8	[Type, File, Java, ByteCode, Void]	9	0.002671
9	[Java, ManifestSection]	9	0.002671
10	[File, Java, Manifest]	9	0.002671
11	[Artifact, File, Jar, Archive, Zip, Java]	9	0.002671
12	[File, Java, ServiceLoader]	10	0.002968
13	[File, Java, Properties]	12	0.003561
14	[Maven, ExecutionGoal]	16	0.004748
15	[Maven, PluginExecution]	16	0.004748
16	[Xml, Attribute]	18	0.005342
17	[jQAssistant, Rule, Concept]	19	0.005638
18	[Maven, Plugin]	21	0.006232
19	[Maven, Configuration]	21	0.006232
20	[Type, File, Java, ByteCode, Enum]	28	0.008309
21	[Type, File, Java, ByteCode, PrimitiveType]	29	0.008606
22	[Git, Branch]	30	0.008903
23	[Xml, Namespace]	36	0.010683
24	[Type, File, Java, ByteCode, Annotation]	43	0.012761
25	[File, Directory]	44	0.013058
26	[Type, File, Java, ByteCode, GenericDeclaratio...	85	0.025225
27	[Artifact, Maven]	92	0.027302
28	[Type, File, Java, ByteCode, ExternalAnnotatio...	93	0.027599
29	[Value, ManifestEntry]	120	0.035611

Chart 1b - Lowest node count by label combination

Shows the lowest (less than 0.5% overall) node count label combinations. Therefore, this plot breaks down the "others" slice of the pie chart above. Values under 0.01% will be grouped into "others" to get a cleaner plot.

<Figure size 640x480 with 0 Axes>

Nodes per label combination (less than 0.5% overall)

Legend:
- ['Xml', 'Text']
- ['Java', 'ByteCode', 'Bound', 'TypeVariable']
- ['Java', 'ByteCode', 'Member', 'Method', 'Lambda']
- ['Type', 'File', 'Java', 'ByteCode', 'ResolvedDuplicateType']
- ['Type', 'File', 'Java', 'ByteCode', 'Class']
- ['Type', 'File', 'Java', 'ByteCode', 'JavaType']
- ['Java', 'Value', 'ByteCode', 'Primitive']
- ['Java', 'ByteCode', 'Member', 'Method', 'GenericDeclaration']
- ['Json', 'Key']
- ['Value', 'Json', 'Scalar']
- ['Type', 'File', 'Java', 'ByteCode', 'ExternalType']
- others
- ['Author', 'Git', 'Person']
- ['Value', 'Array']
- ['Committer', 'Git', 'Person']
- ['Type', 'File', 'Java', 'ByteCode', 'Class', 'GenericDeclaration']
- ['Value', 'Property']
- ['Type', 'File', 'Java', 'ByteCode', 'Interface']
- ['Java', 'Value', 'ByteCode', 'Class']
- ['Git', 'Tag']
- ['Maven', 'Dependency']
- ['Java', 'ByteCode', 'Bound', 'GenericArrayType']
- ['Java', 'Value', 'ByteCode', 'Enum']
- ['Package', 'File', 'Directory', 'Java']
- ['Value', 'Json', 'Object']
- ['Value', 'ManifestEntry']
- ['Type', 'File', 'Java', 'ByteCode', 'ExternalAnnotation', 'ExternalType']
- ['Artifact', 'Maven']
- ['Type', 'File', 'Java', 'ByteCode', 'GenericDeclaration', 'Interface']
- ['File', 'Directory']
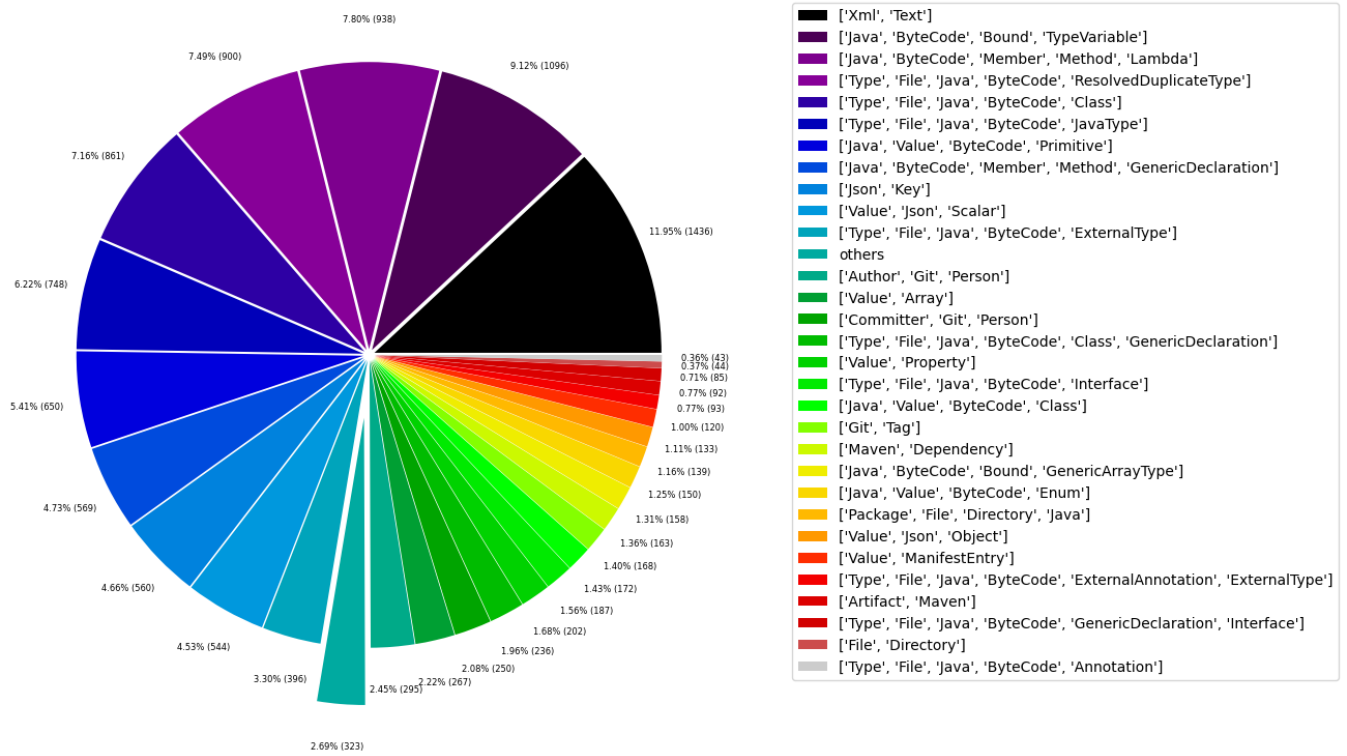- ['Type', 'File', 'Java', 'ByteCode', 'Annotation']

Table 1c - Highest node count by single label

Lists the 40 labels with the highest number of nodes. Doesn't sum up to the total number of nodes or 100% because one node can have multiple labels. Helps to identify commonly used labels.
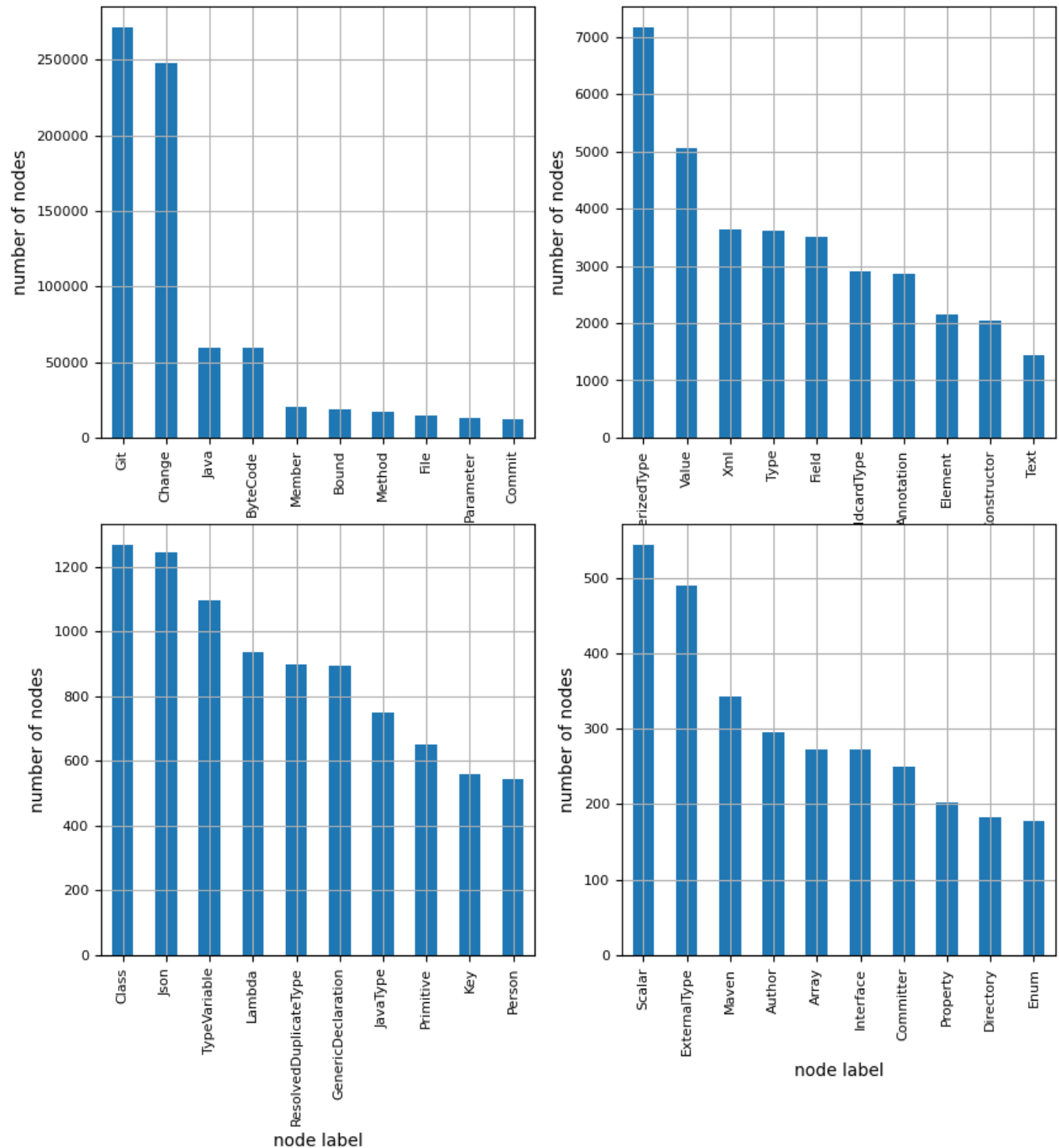
	nodeLabel	nodesWithThatLabel	nodesWithThatLabelPercent
0	Git	271775	80.652341
1	Change	248143	73.639275
2	Java	59318	17.603295
3	ByteCode	59130	17.547504
4	Member	20258	6.011793
5	Bound	18447	5.474358
6	Method	16755	4.972238
7	File	14547	4.316989
8	Parameter	13016	3.862647
9	Commit	12194	3.618709
10	ParameterizedType	7172	2.128373
11	Value	5066	1.503393
12	Xml	3643	1.081102
13	Type	3615	1.072793
14	Field	3503	1.039555
15	WildcardType	2908	0.862982
16	Annotation	2865	0.850222
17	Element	2144	0.636257
18	Constructor	2047	0.607471
19	Text	1436	0.426149
20	Class	1269	0.376590
21	Json	1245	0.369468
22	TypeVariable	1096	0.325251
23	Lambda	938	0.278362
24	ResolvedDuplicateType	900	0.267085
25	GenericDeclaration	894	0.265305
26	JavaType	748	0.221978
27	Primitive	650	0.192895
28	Key	560	0.166186
29	Person	545	0.161735
30	Scalar	544	0.161438
31	ExternalType	489	0.145116
32	Maven	343	0.101789
33	Author	295	0.087545
34	Array	273	0.081016
35	Interface	272	0.080719
36	Committer	250	0.074190
37	Property	202	0.059946
38	Directory	183	0.054307
39	Enum	178	0.052824

Chart 1c - Highest node count by label

Shows the 40 labels with the highest number of nodes.

<Figure size 640x480 with 0 Axes>

Node count by label

Relationship Types

Table 2a - Highest relationship count by type

Lists the 30 relationship types with the highest number of occurrences. The whole table can be found in the CSV report `Relationship_type_count`.

```
Total number of relationships: 1045278
```

	relationshipType	nodesWithThatRelationshipType	nodesWithThatRelationshipTypePercent
0	CONTAINS_CHANGE	248143	23.739426
1	MODIFIES	248143	23.739426
2	UPDATES	161433	15.444025
3	CREATES	59774	5.718479
4	DELETES	37226	3.561349
5	INVOKES	35680	3.413446
6	COMMITTED	24388	2.333159
7	DEPENDS_ON	21887	2.093893
8	OF_TYPE	21373	2.044719
9	DECLARES	20699	1.980239
10	OF_RAW_TYPE	17084	1.634398
11	HAS_PARENT	14816	1.417422
12	HAS	14110	1.349880
13	RETURNS	12562	1.201786
14	HAS_COMMIT	12194	1.166580
15	HAS_FILE	10694	1.023077
16	RENAMES	10290	0.984427
17	READS	9097	0.870295
18	HAS_ACTUAL_TYPE_ARGUMENT	8284	0.792516
19	HAS_NEW_NAME	6199	0.593048
20	OF_GENERIC_TYPE	5903	0.564730
21	RESOLVES_TO	5206	0.498049
22	SIMILAR	3968	0.379612
23	WRITES	3812	0.364688
24	CONTAINS	3806	0.364114
25	RETURNS_GENERIC	3543	0.338953
26	ANNOTATED_BY	2810	0.268828
27	REQUIRES	2175	0.208079
28	HAS_FIRST_CHILD	2144	0.205113
29	HAS_LAST_CHILD	2144	0.205113

Chart 2a - Highest relationship count by type

Values under 0.5% will be grouped into "others" to get a cleaner plot. The group "others" is then broken down in the second chart.

```
<Figure size 640x480 with 0 Axes>
```
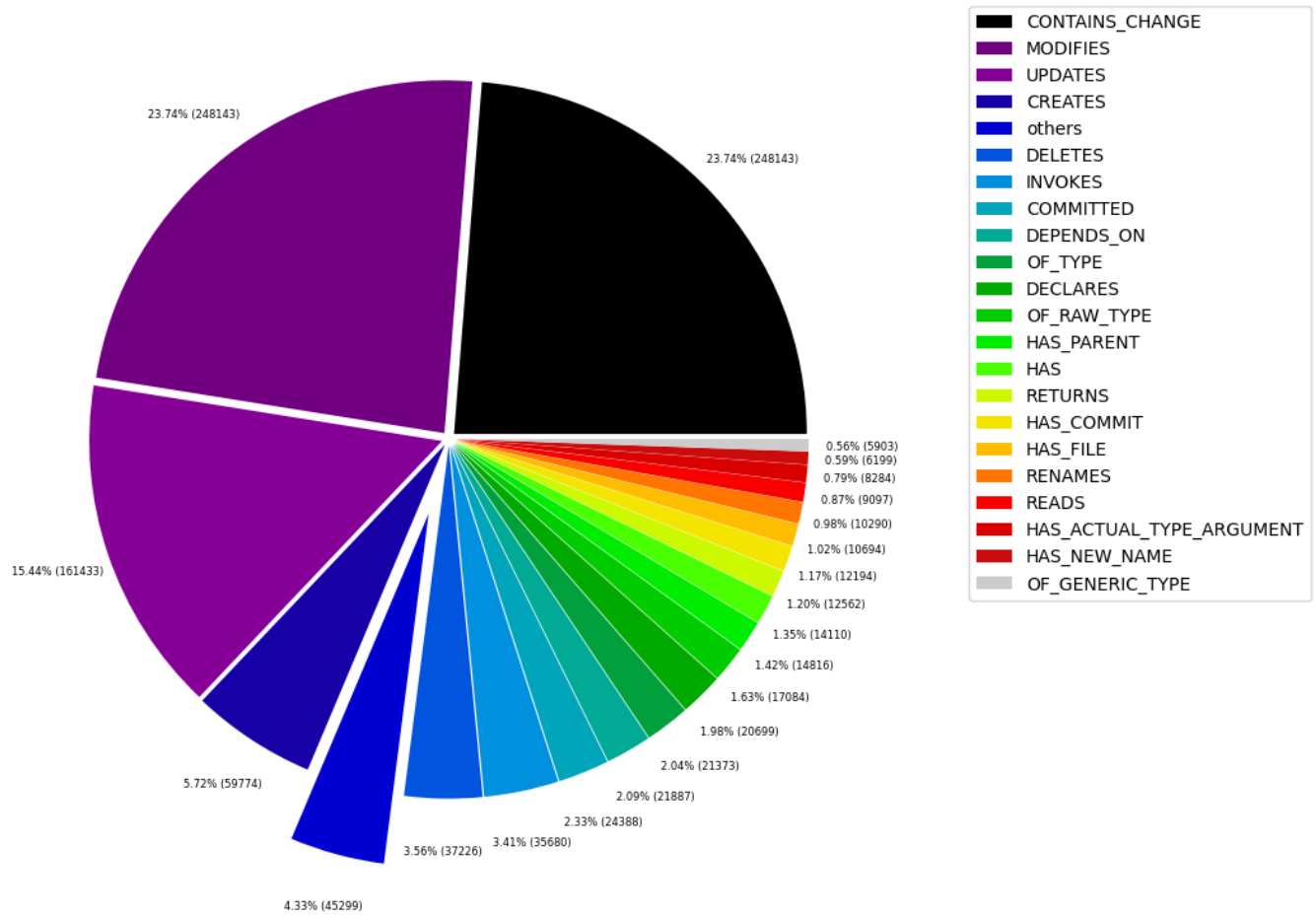
Relationship types (more than 0.5% overall)

Legend:
- CONTAINS_CHANGE
- MODIFIES
- UPDATES
- CREATES
- others
- DELETES
- INVOKES
- COMMITTED
- DEPENDS_ON
- OF_TYPE
- DECLARES
- OF_RAW_TYPE
- HAS_PARENT
- HAS
- RETURNS
- HAS_COMMIT
- HAS_FILE
- RENAMES
- READS
- HAS_ACTUAL_TYPE_ARGUMENT
- HAS_NEW_NAME
- OF_GENERIC_TYPE

Slice values:
- 23.74% (248143)
- 23.74% (248143)
- 15.44% (161433)
- 5.72% (59774)
- 4.33% (45299)
- 3.56% (37226)
- 3.41% (35680)
- 2.33% (24388)
- 2.09% (21887)
- 2.04% (21373)
- 1.98% (20699)
- 1.63% (17084)
- 1.42% (14816)
- 1.35% (14110)
- 1.20% (12562)
- 1.17% (12194)
- 1.02% (10694)
- 0.98% (10290)
- 0.87% (9097)
- 0.79% (8284)
- 0.59% (6199)
- 0.56% (5903)

Table 2b - Lowest relationship count by type

Lists the 30 relationships type with the lowest number of occurrences up to 0.5% of the total node count. This is essentially breaking down the "others" slice from the chart above.
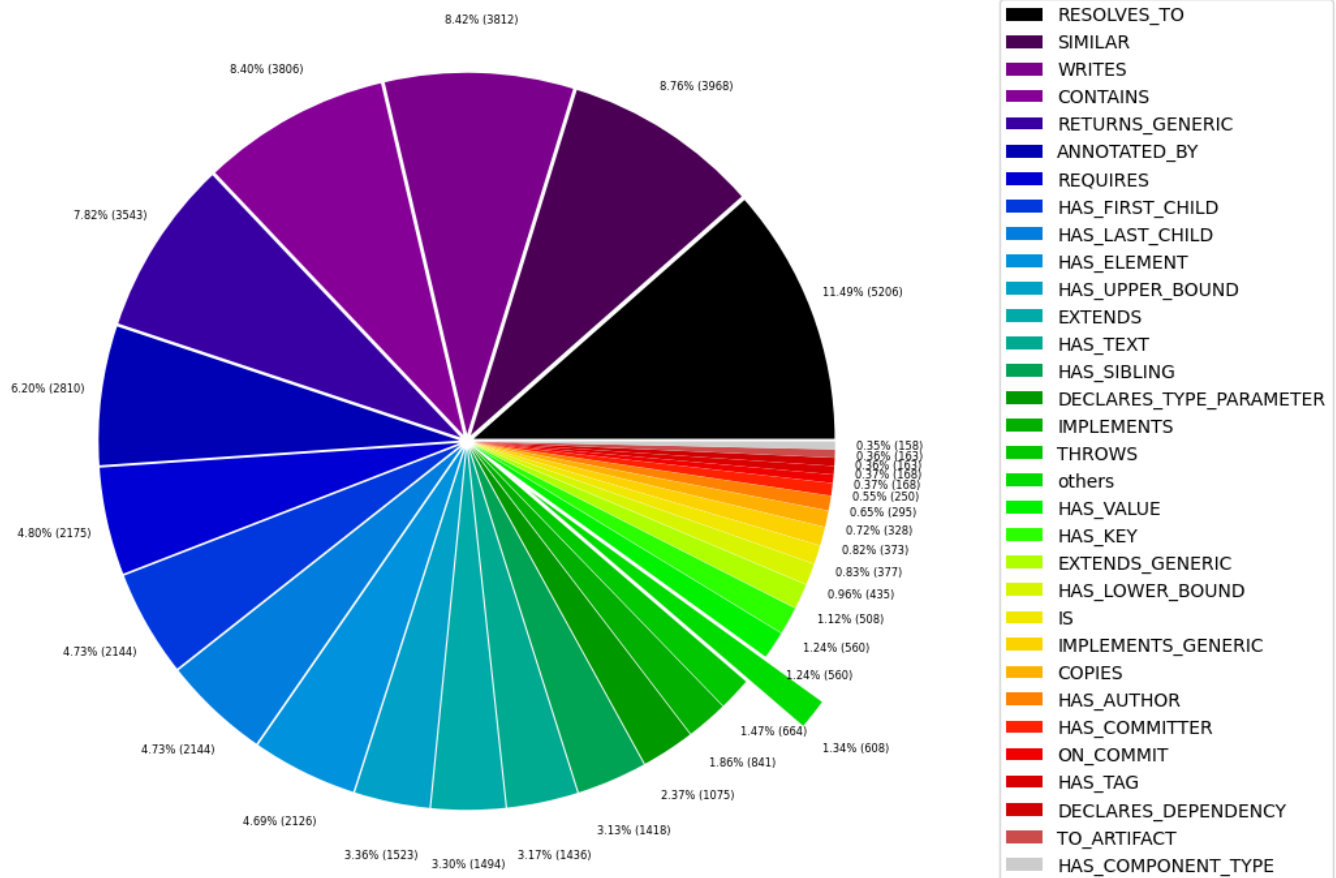
	relationshipType	nodesWithThatRelationshipType	nodesWithThatRelationshipTypePercent
0	HAS_PROPERTY	1	0.000096
1	THROWS_GENERIC	5	0.000478
2	EXCLUDES	5	0.000478
3	HAS_ROOT_ELEMENT	9	0.000861
4	DESCRIBES	9	0.000861
5	HAS_GOAL	16	0.001531
6	HAS_EXECUTION	16	0.001531
7	OF_NAMESPACE	18	0.001722
8	HAS_ATTRIBUTE	18	0.001722
9	INCLUDES_CONCEPT	19	0.001818
10	USES_PLUGIN	21	0.002009
11	REQUIRES_TYPE_PARAMETER	21	0.002009
12	IS_ARTIFACT	21	0.002009
13	HAS_CONFIGURATION	21	0.002009
14	REQUIRES_CONCEPT	28	0.002679
15	HAS_BRANCH	30	0.002870
16	HAS_HEAD	31	0.002966
17	DECLARES_NAMESPACE	36	0.003444
18	HAS_DEFAULT	36	0.003444
19	CONTAINS_VALUE	121	0.011576
20	COPY_OF	126	0.012054
21	HAS_COMPONENT_TYPE	158	0.015116
22	TO_ARTIFACT	163	0.015594
23	DECLARES_DEPENDENCY	163	0.015594
24	ON_COMMIT	168	0.016072
25	HAS_TAG	168	0.016072
26	HAS_COMMITTER	250	0.023917
27	HAS_AUTHOR	295	0.028222
28	COPIES	328	0.031379
29	IMPLEMENTS_GENERIC	373	0.035684

Chart 2b - Lowest relationship count by type

Shows the lowest (less than 0.5% overall) relationship types. This plot breaks down the "others" slice of the pie chart above. Values under 0.01% will be grouped into "others" to get a cleaner plot.

```
<Figure size 640x480 with 0 Axes>
```

Relationship types (less than 0.5% overall)

Legend
RESOLVES_TO
SIMILAR
WRITES
CONTAINS
RETURNS_GENERIC
ANNOTATED_BY
REQUIRES
HAS_FIRST_CHILD
HAS_LAST_CHILD
HAS_ELEMENT
HAS_UPPER_BOUND
EXTENDS
HAS_TEXT
HAS_SIBLING
DECLARES_TYPE_PARAMETER
IMPLEMENTS
THROWS
others
HAS_VALUE
HAS_KEY
EXTENDS_GENERIC
HAS_LOWER_BOUND
IS
IMPLEMENTS_GENERIC
COPIES
HAS_AUTHOR
HAS_COMMITTER
ON_COMMIT
HAS_TAG
DECLARES_DEPENDENCY
TO_ARTIFACT
HAS_COMPONENT_TYPE

Node labels with their relationships

Table 3a - Highest relationship count by node labels and relationship type

Lists the 30 node labels and their relationship types with the highest number of occurrences.

	sourceLabels	relationType	targetLabels	numberOfRelationships	numberOfNodesWithSameLabelsAsSource	numberOfNodesWithS
0	[Git, Commit]	CONTAINS_CHANGE	[Git, Change]	248143	12194	
1	[Git, Change]	MODIFIES	[File, Git]	248143	248143	
2	[Git, Change]	UPDATES	[File, Git]	161433	248143	
3	[Git, Change]	CREATES	[File, Git]	59774	248143	
4	[Git, Change]	DELETES	[File, Git]	37226	248143	
5	[Java, ByteCode, Member, Method]	INVOKES	[Java, ByteCode, Member, Method]	21558	13075	
6	[Git, Commit]	HAS_PARENT	[Git, Commit]	14807	12194	
7	[Repository, File, Git]	HAS_COMMIT	[Git, Commit]	12194	1	
8	[Author, Git, Person]	COMMITTED	[Git, Commit]	12194	295	
9	[Committer, Git, Person]	COMMITTED	[Git, Commit]	12194	250	
10	[Repository, File, Git]	HAS_FILE	[File, Git]	10694	1	
11	[Git, Change]	RENAMES	[File, Git]	10290	248143	
12	[Java, ByteCode, Member, Method]	HAS	[Java, ByteCode, Parameter]	8320	13075	
13	[Java, ByteCode, Member, Method]	READS	[Java, ByteCode, Member, Field]	8121	13075	
14	[Java, ByteCode, Parameter]	OF_TYPE	[Type, File, Java, ByteCode, JavaType]	6368	13016	
15	[File, Git]	HAS_NEW_NAME	[File, Git]	6199	10694	
16	[Java, ByteCode, Bound]	OF_RAW_TYPE	[Type, File, Java, ByteCode, JavaType]	3570	7113	
17	[Java, ByteCode, ParameterizedType, Bound]	OF_RAW_TYPE	[Type, File, Java, ByteCode, JavaType]	3049	7172	
18	[Java, ByteCode, ParameterizedType, Bound]	HAS_ACTUAL_TYPE_ARGUMENT	[Java, ByteCode, Bound, WildcardType]	2908	7172	
19	[Java, ByteCode, Parameter]	OF_GENERIC_TYPE	[Java, ByteCode, ParameterizedType, Bound]	2655	13016	
20	[Java, ByteCode, ParameterizedType, Bound]	HAS_ACTUAL_TYPE_ARGUMENT	[Java, ByteCode, Bound, TypeVariable]	2454	7172	
21	[Java, ByteCode, Member, Constructor, Method]	WRITES	[Java, ByteCode, Member, Field]	2432	2043	
22	[Java, Value, ByteCode, Annotation]	OF_TYPE	[Type, File, Java, ByteCode, ExternalAnnotatio...]	2341	2822	
23	[Xml, Element]	HAS_ELEMENT	[Xml, Element]	2126	2144	
24	[Java, ByteCode, Member, Constructor, Method]	INVOKES	[Java, ByteCode, Member, Constructor, Method]	2113	2043	
25	[Java, ByteCode, Member, Method]	RETURNS	[Type, File, Java, ByteCode, JavaType]	2088	13075	
26	[Java, ByteCode, Member, Method]	INVOKES	[Java, ByteCode, Member, Constructor, Method]	2030	13075	
27	[Java, ByteCode, Member, Constructor, Method]	HAS	[Java, ByteCode, Parameter]	2023	2043	
28	[Java, ByteCode, Parameter]	OF_GENERIC_TYPE	[Java, ByteCode, Bound]	2023	13016	
29	[Java, ByteCode, Member, Method, Lambda]	INVOKES	[Java, ByteCode, Member, Method]	1879	938	

Graph Density

```
total_number_of_nodes (vertices): 336971
total_number_of_relationships (edges): 1045278
-> total directed graph density: 9.205514050727518e-06
-> total directed graph density in percent: 0.0009205514050727517
```