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]	249236	73.708150
1	[Java, ByteCode, Member, Method]	13202	3.904312
2	[Java, ByteCode, Parameter]	13029	3.853149
3	[Git, Commit]	12227	3.615969
4	[File, Git]	10703	3.165266
5	[Java, ByteCode, ParameterizedType, Bound]	7176	2.122204
6	[Java, ByteCode, Bound]	7118	2.105051
7	[Java, ByteCode, Member, Field]	3504	1.036260
8	[Java, ByteCode, Bound, WildcardType]	2908	0.860001
9	[Java, Value, ByteCode, Annotation]	2822	0.834568
10	[Xml, Element]	2144	0.634059
11	[Java, ByteCode, Member, Constructor, Method]	2042	0.603894
12	[Xml, Text]	1436	0.424677
13	[Java, ByteCode, Bound, TypeVariable]	1096	0.324127
14	[Java, ByteCode, Member, Method, Lambda]	942	0.278584
15	[Type, File, Java, ByteCode, ResolvedDuplicate...	899	0.265867
16	[Type, File, Java, ByteCode, Class]	864	0.255516
17	[Type, File, Java, ByteCode, JavaType]	748	0.221211
18	[Java, Value, ByteCode, Primitive]	652	0.192820
19	[Java, ByteCode, Member, Method, GenericDeclar...	569	0.168274
20	[Json, Key]	560	0.165612
21	[Value, Json, Scalar]	544	0.160881
22	[Type, File, Java, ByteCode, ExternalType]	396	0.117112
23	[Author, Git, Person]	295	0.087242
24	[Value, Array]	267	0.078962
25	[Committer, Git, Person]	250	0.073934
26	[Type, File, Java, ByteCode, Class, GenericDec...	236	0.069794
27	[Value, Property]	202	0.059739
28	[Type, File, Java, ByteCode, Interface]	188	0.055598
29	[Java, Value, ByteCode, Class]	172	0.050867

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)

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']

73.71% (249236)
0.60% (2042)
0.63% (2144)
0.83% (2822)
0.86% (2908)
1.04% (3504)
2.11% (7118)
2.12% (7176)
3.17% (10703)
3.56% (12028)
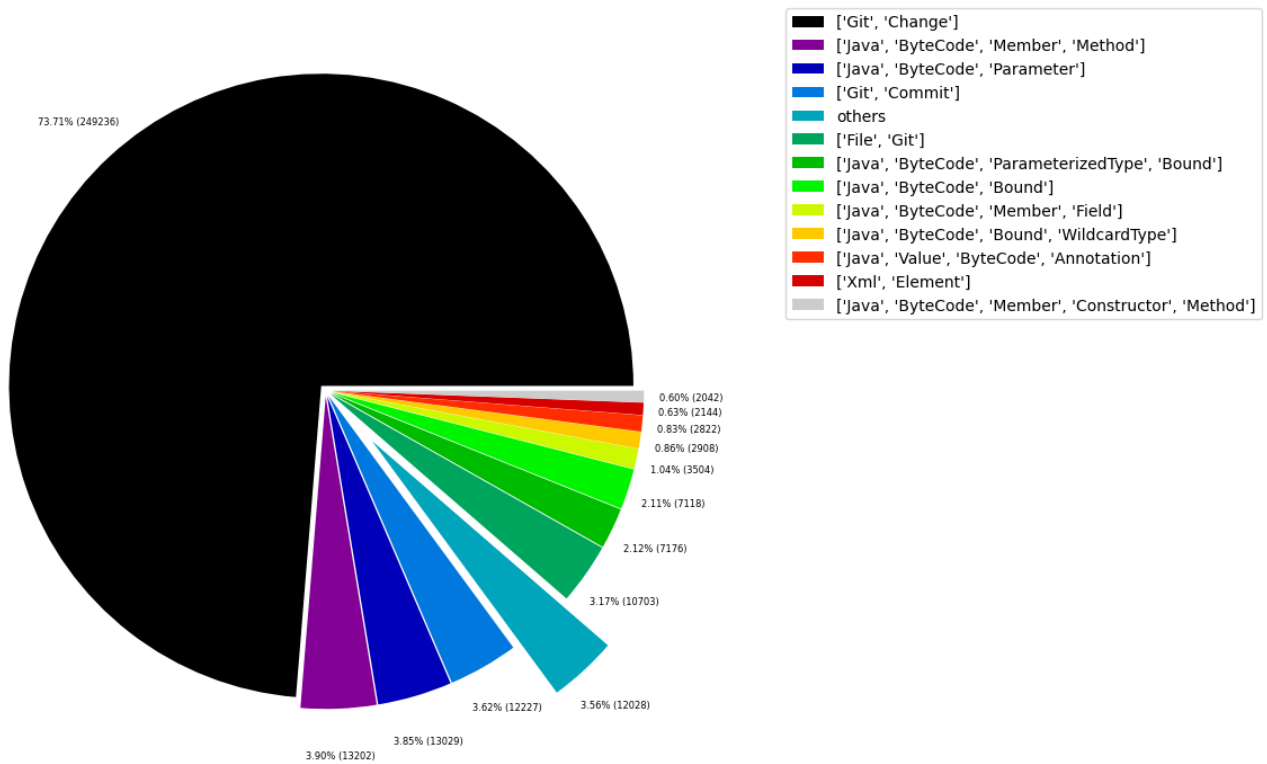3.62% (12227)
3.85% (13029)
3.90% (13202)

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	0.000296
1	[Repository, File, Git]	1	0.000296
2	[File, Json]	2	0.000591
3	[File]	3	0.000887
4	[Java, ByteCode, Member, Constructor, Method, ...	4	0.001183
5	[Maven, Exclusion]	5	0.001479
6	[Value, Json, Array]	6	0.001774
7	[File, Maven, Xml, Pom, Document]	9	0.002662
8	[Type, File, Java, ByteCode, Void]	9	0.002662
9	[Java, ManifestSection]	9	0.002662
10	[File, Java, Manifest]	9	0.002662
11	[Artifact, File, Jar, Archive, Zip, Java]	9	0.002662
12	[File, Java, ServiceLoader]	10	0.002957
13	[File, Java, Properties]	12	0.003549
14	[Maven, ExecutionGoal]	16	0.004732
15	[Maven, PluginExecution]	16	0.004732
16	[Xml, Attribute]	18	0.005323
17	[jQAssistant, Rule, Concept]	19	0.005619
18	[Maven, Plugin]	21	0.006210
19	[Maven, Configuration]	21	0.006210
20	[Type, File, Java, ByteCode, Enum]	28	0.008281
21	[Type, File, Java, ByteCode, PrimitiveType]	29	0.008576
22	[Git, Branch]	30	0.008872
23	[Xml, Namespace]	36	0.010647
24	[Type, File, Java, ByteCode, Annotation]	43	0.012717
25	[File, Directory]	44	0.013012
26	[Type, File, Java, ByteCode, GenericDeclaratio...	85	0.025138
27	[Artifact, Maven]	92	0.027208
28	[Type, File, Java, ByteCode, ExternalAnnotatio...	93	0.027503
29	[Value, ManifestEntry]	120	0.035488

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']
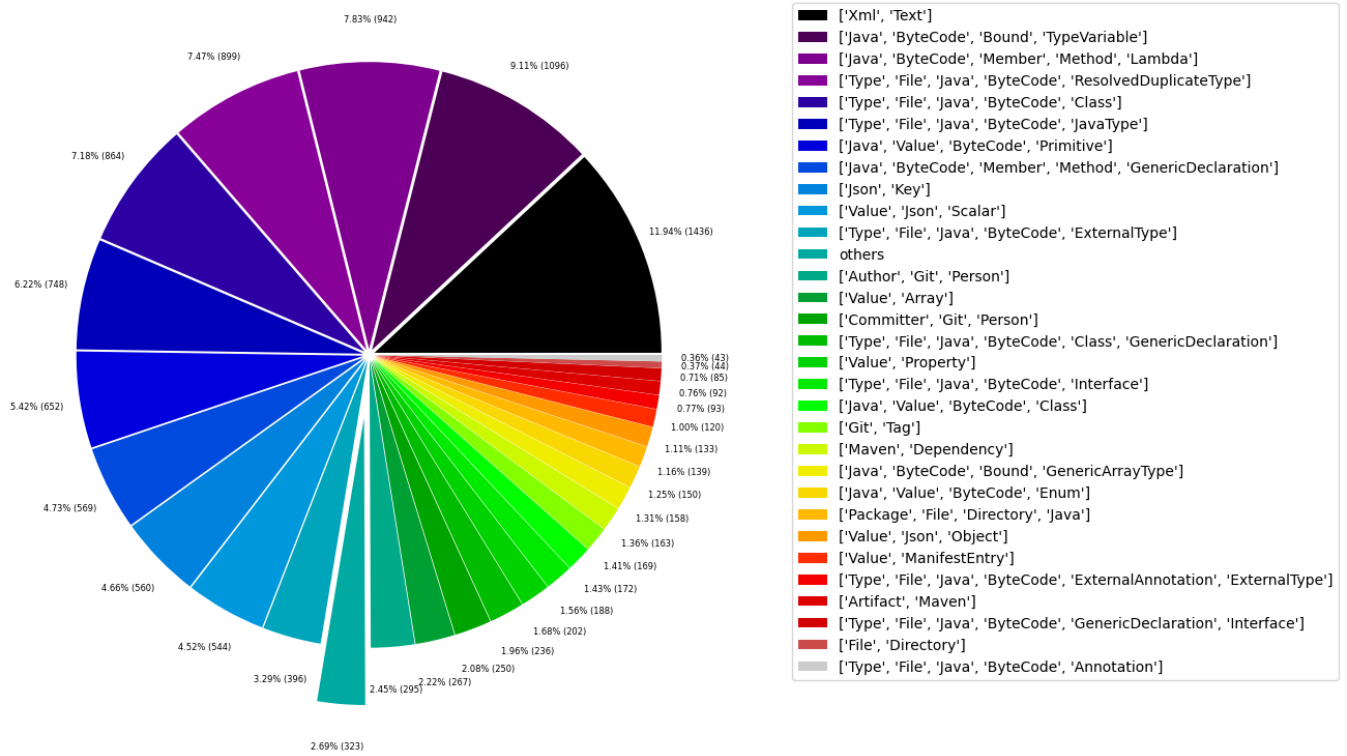- ['Type', 'File', 'Java', 'ByteCode', 'Annotation']

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.
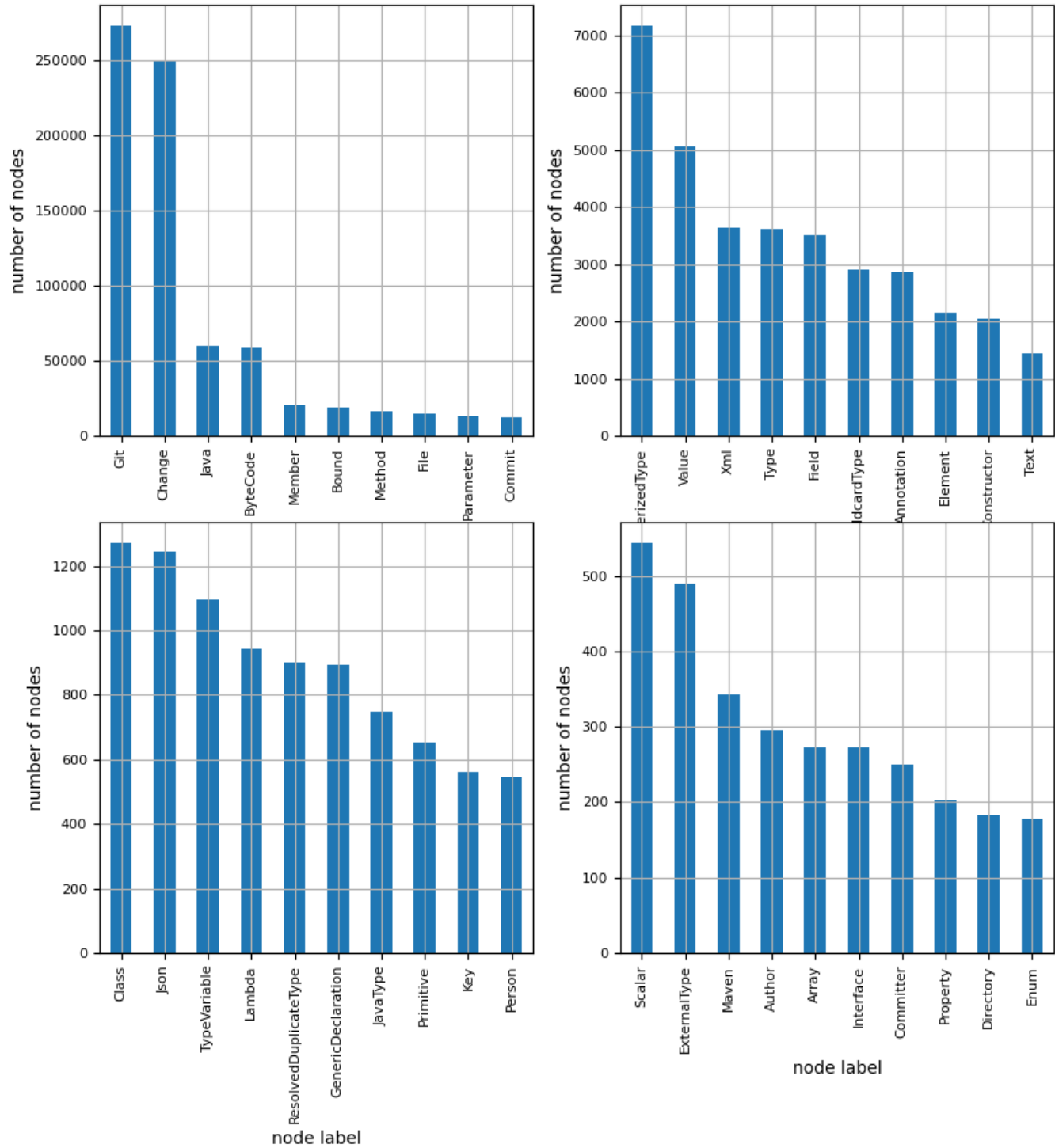
	nodeLabel	nodesWithThatLabel	nodesWithThatLabelPercent
0	Git	272911	80.709708
1	Change	249236	73.708150
2	Java	59350	17.551953
3	ByteCode	59162	17.496355
4	Member	20263	5.992506
5	Bound	18456	5.458110
6	Method	16759	4.956246
7	File	14559	4.305626
8	Parameter	13029	3.853149
9	Commit	12227	3.615969
10	ParameterizedType	7176	2.122204
11	Value	5068	1.498792
12	Xml	3643	1.077368
13	Type	3618	1.069974
14	Field	3504	1.036260
15	WildcardType	2908	0.860001
16	Annotation	2865	0.847285
17	Element	2144	0.634059
18	Constructor	2046	0.605077
19	Text	1436	0.424677
20	Class	1272	0.376177
21	Json	1245	0.368192
22	TypeVariable	1096	0.324127
23	Lambda	942	0.278584
24	ResolvedDuplicateType	899	0.265867
25	GenericDeclaration	894	0.264388
26	JavaType	748	0.221211
27	Primitive	652	0.192820
28	Key	560	0.165612
29	Person	545	0.161176
30	Scalar	544	0.160881
31	ExternalType	489	0.144615
32	Maven	343	0.101438
33	Author	295	0.087242
34	Array	273	0.080736
35	Interface	273	0.080736
36	Committer	250	0.073934
37	Property	202	0.059739
38	Directory	183	0.054120
39	Enum	178	0.052641

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: 1048991
```

	relationshipType	nodesWithThatRelationshipType	nodesWithThatRelationshipTypePercent
0	CONTAINS_CHANGE	249236	23.759594
1	MODIFIES	249236	23.759594
2	UPDATES	162500	15.491077
3	CREATES	59791	5.699858
4	DELETES	37236	3.549697
5	INVOKES	35733	3.406416
6	COMMITTED	24454	2.331193
7	DEPENDS_ON	21931	2.090676
8	OF_TYPE	21394	2.039484
9	DECLARES	20708	1.974087
10	OF_RAW_TYPE	17093	1.629471
11	HAS_PARENT	14859	1.416504
12	HAS	14125	1.346532
13	RETURNS	12578	1.199057
14	HAS_COMMIT	12227	1.165596
15	HAS_FILE	10703	1.020314
16	RENAMES	10291	0.981038
17	READS	9161	0.873315
18	HAS_ACTUAL_TYPE_ARGUMENT	8288	0.790093
19	HAS_NEW_NAME	6200	0.591044
20	OF_GENERIC_TYPE	5906	0.563017
21	RESOLVES_TO	5215	0.497144
22	SIMILAR	3977	0.379126
23	WRITES	3818	0.363969
24	CONTAINS	3814	0.363587
25	RETURNS_GENERIC	3545	0.337944
26	ANNOTATED_BY	2810	0.267876
27	REQUIRES	2174	0.207247
28	HAS_FIRST_CHILD	2144	0.204387
29	HAS_LAST_CHILD	2144	0.204387

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>
```
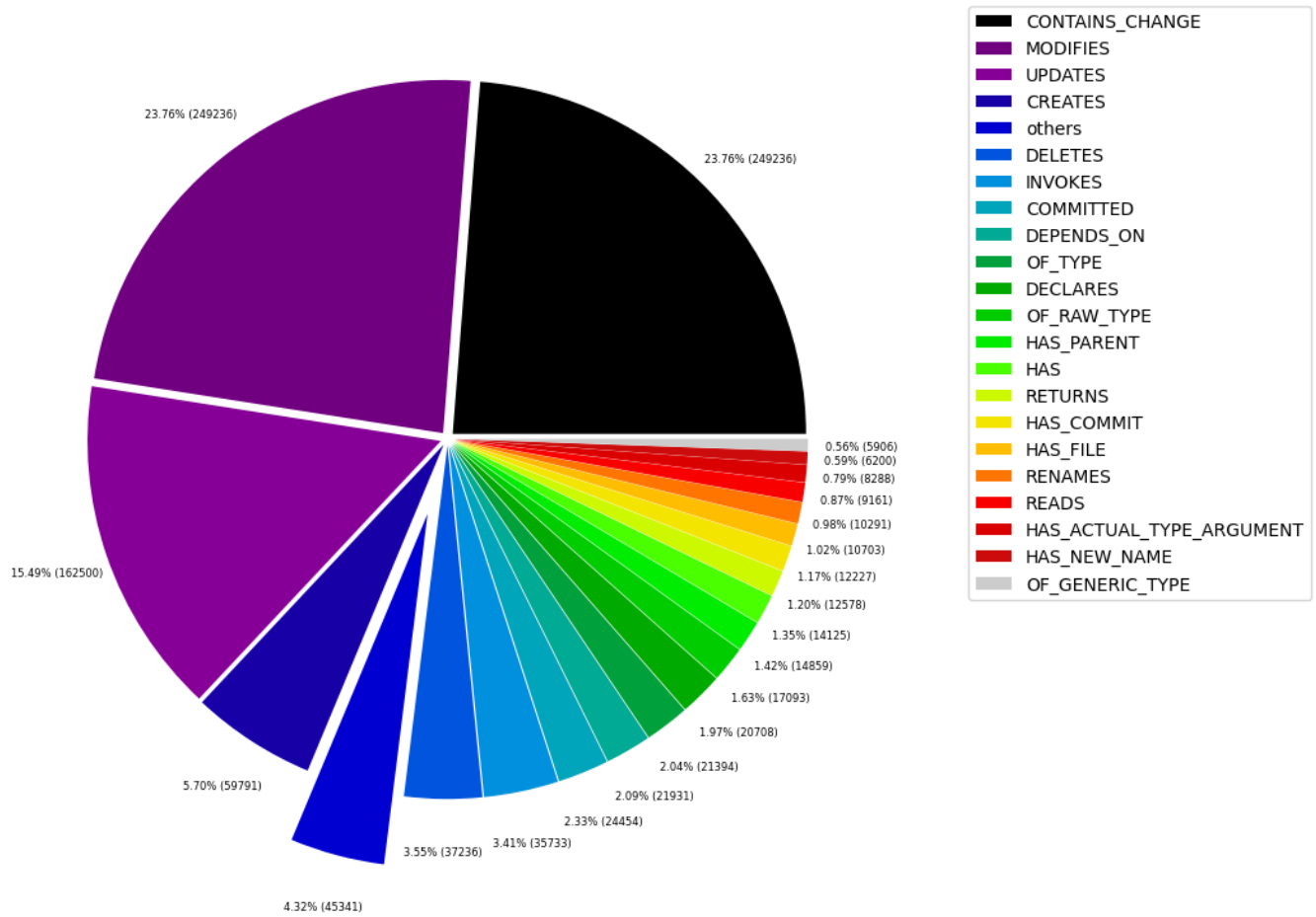
Relationship types (more than 0.5% overall)

23.76% (249236)

23.76% (249236)

15.49% (162500)

5.70% (59791)

3.55% (37236)

3.41% (35733)

2.33% (24454)

4.32% (45341)

2.09% (21931)

2.04% (21394)

1.97% (20708)

1.63% (17093)

1.42% (14859)

1.35% (14125)

1.20% (12578)

1.17% (12227)

1.02% (10703)

0.98% (10291)

0.87% (9161)

0.79% (8288)

0.59% (6200)

0.56% (5906)

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

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.
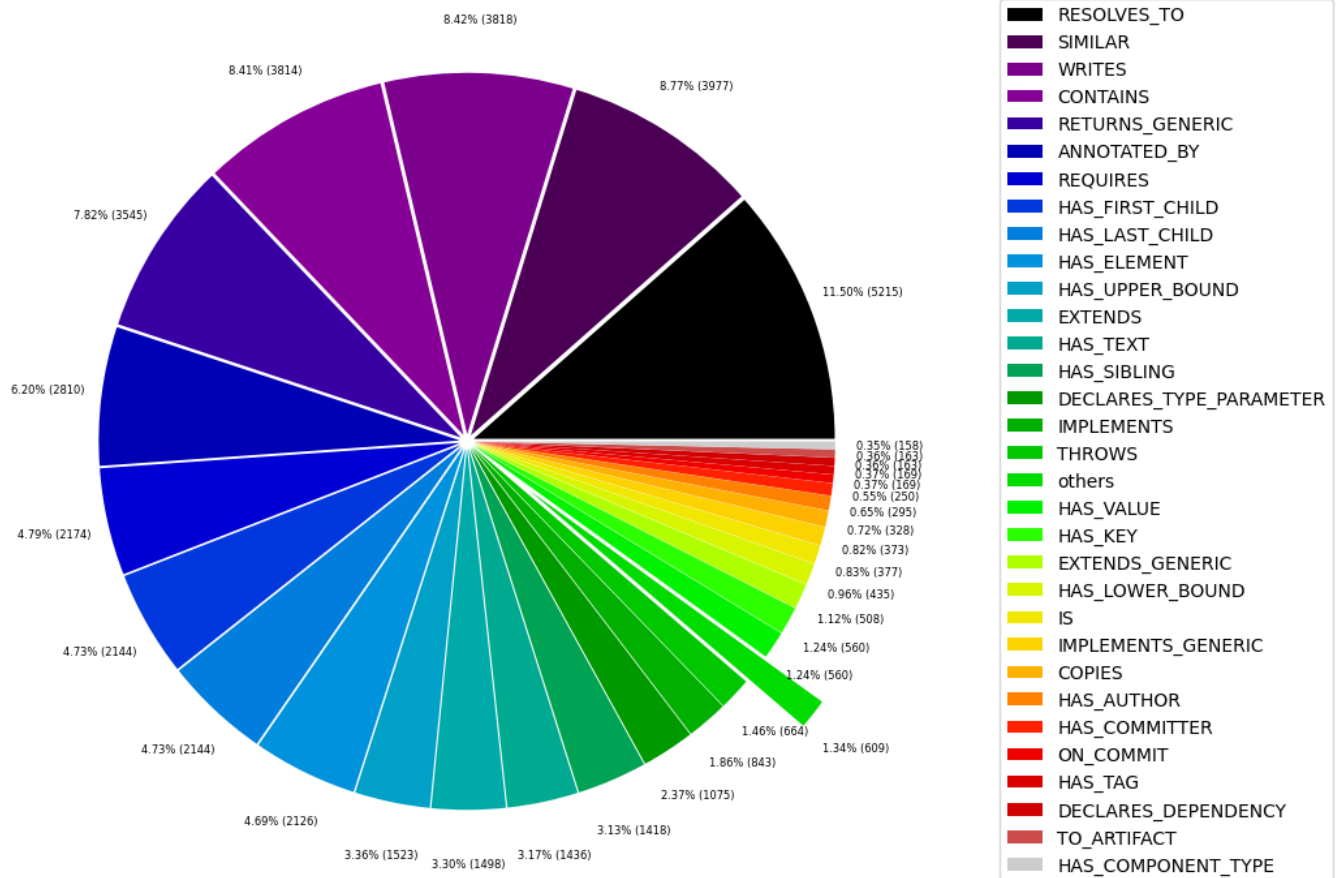
	relationshipType	nodesWithThatRelationshipType	nodesWithThatRelationshipTypePercent
0	HAS_PROPERTY	1	0.000095
1	THROWS_GENERIC	5	0.000477
2	EXCLUDES	5	0.000477
3	DESCRIBES	9	0.000858
4	HAS_ROOT_ELEMENT	10	0.000953
5	HAS_GOAL	16	0.001525
6	HAS_EXECUTION	16	0.001525
7	OF_NAMESPACE	18	0.001716
8	HAS_ATTRIBUTE	18	0.001716
9	INCLUDES_CONCEPT	19	0.001811
10	USES_PLUGIN	21	0.002002
11	REQUIRES_TYPE_PARAMETER	21	0.002002
12	IS_ARTIFACT	21	0.002002
13	HAS_CONFIGURATION	21	0.002002
14	REQUIRES_CONCEPT	28	0.002669
15	HAS_BRANCH	30	0.002860
16	HAS_HEAD	31	0.002955
17	DECLARES_NAMESPACE	36	0.003432
18	HAS_DEFAULT	36	0.003432
19	CONTAINS_VALUE	121	0.011535
20	COPY_OF	126	0.012012
21	HAS_COMPONENT_TYPE	158	0.015062
22	TO_ARTIFACT	163	0.015539
23	DECLARES_DEPENDENCY	163	0.015539
24	ON_COMMIT	169	0.016111
25	HAS_TAG	169	0.016111
26	HAS_COMMITTER	250	0.023832
27	HAS_AUTHOR	295	0.028122
28	COPIES	328	0.031268
29	IMPLEMENTS_GENERIC	373	0.035558

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]	249236	12227	
1	[Git, Change]	MODIFIES	[File, Git]	249236	249236	
2	[Git, Change]	UPDATES	[File, Git]	162500	249236	
3	[Git, Change]	CREATES	[File, Git]	59791	249236	
4	[Git, Change]	DELETES	[File, Git]	37236	249236	
5	[Java, ByteCode, Member, Method]	INVOKES	[Java, ByteCode, Member, Method]	21635	13105	
6	[Git, Commit]	HAS_PARENT	[Git, Commit]	14850	12227	
7	[Repository, File, Git]	HAS_COMMIT	[Git, Commit]	12227	1	
8	[Author, Git, Person]	COMMITTED	[Git, Commit]	12227	295	
9	[Committer, Git, Person]	COMMITTED	[Git, Commit]	12227	250	
10	[Repository, File, Git]	HAS_FILE	[File, Git]	10703	1	
11	[Git, Change]	RENAMES	[File, Git]	10291	249236	
12	[Java, ByteCode, Member, Method]	HAS	[Java, ByteCode, Parameter]	8357	13105	
13	[Java, ByteCode, Member, Method]	READS	[Java, ByteCode, Member, Field]	8184	13105	
14	[Java, ByteCode, Parameter]	OF_TYPE	[Type, File, Java, ByteCode, JavaType]	6371	13029	
15	[File, Git]	HAS_NEW_NAME	[File, Git]	6200	10703	
16	[Type, File, Java, ByteCode, Class]	DECLARES	[Java, ByteCode, Member, Method]	5089	816	
17	[Type, File, Java, ByteCode, Class]	DEPENDS_ON	[Type, File, Java, ByteCode, JavaType]	4493	816	
18	[Java, ByteCode, Bound]	OF_RAW_TYPE	[Type, File, Java, ByteCode, JavaType]	3570	7118	
19	[Java, ByteCode, ParameterizedType, Bound]	OF_RAW_TYPE	[Type, File, Java, ByteCode, JavaType]	3053	7176	
20	[Java, ByteCode, ParameterizedType, Bound]	HAS_ACTUAL_TYPE_ARGUMENT	[Java, ByteCode, Bound, WildcardType]	2908	7176	
21	[Java, ByteCode, Parameter]	OF_GENERIC_TYPE	[Java, ByteCode, ParameterizedType, Bound]	2657	13029	
22	[Java, ByteCode, ParameterizedType, Bound]	HAS_ACTUAL_TYPE_ARGUMENT	[Java, ByteCode, Bound, TypeVariable]	2454	7176	
23	[Java, ByteCode, Member, Constructor, Method]	WRITES	[Java, ByteCode, Member, Field]	2438	2042	
24	[Java, Value, ByteCode, Annotation]	OF_TYPE	[Type, File, Java, ByteCode, ExternalAnnotatio...]	2341	2822	
25	[Type, File, Java, ByteCode, Class]	DECLARES	[Java, ByteCode, Member, Field]	2184	816	
26	[Xml, Element]	HAS_ELEMENT	[Xml, Element]	2126	2144	
27	[Java, ByteCode, Member, Constructor, Method]	INVOKES	[Java, ByteCode, Member, Constructor, Method]	2120	2042	
28	[Type, File, Java, ByteCode, Class, GenericDec...]	DECLARES	[Java, ByteCode, Member, Method]	2093	228	
29	[Java, ByteCode, Member, Method]	RETURNS	[Type, File, Java, ByteCode, JavaType]	2091	13105	

Graph Density

```
total_number_of_nodes (vertices): 338139
total_number_of_relationships (edges): 1048991
-> total directed graph density: 9.174502398840212e-06
-> total directed graph density in percent: 0.0009174502398840213
```