Bayesian Modeling for Software Quality

By

Mohammad Sharif Hanif
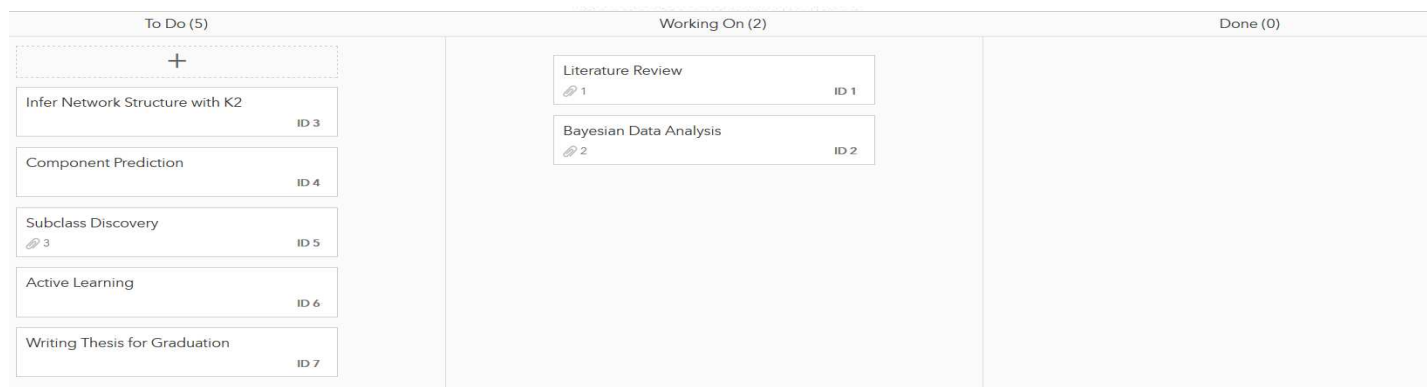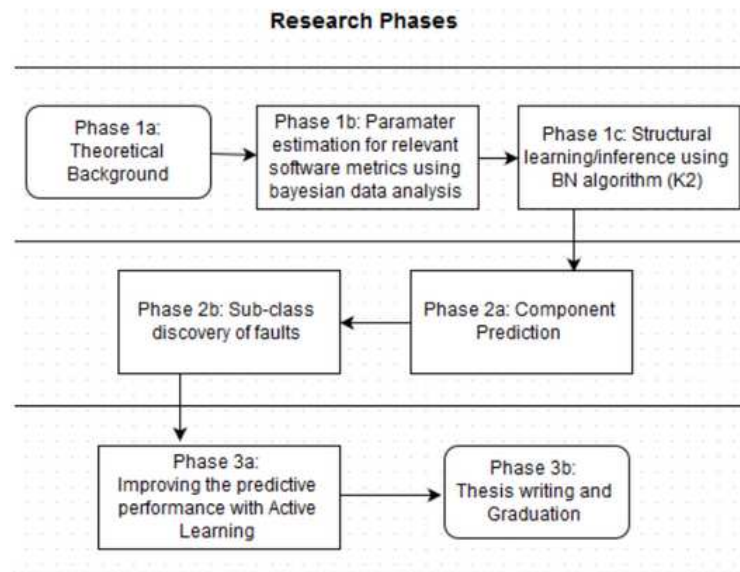
Department of Computer Science

Brunel University London

30 March 2015

Today's agenda

- PhD plan
- Research Progress
- Next Step
- Project timetable
- Admin matters
- Next meeting

Research Phases

To Do (5)	Working On (2)	Done (0)
+	Literature Review 🔗 1 ID 1	
Infer Network Structure with K2 ID 3	Bayesian Data Analysis 🔗 2 ID 2	
Component Prediction ID 4		
Subclass Discovery 🔗 3 ID 5		
Active Learning ID 6		
Writing Thesis for Graduation ID 7		

defect-reports.arff

Relation: defect-reports

No. Numeric	id Numeric	description String	priority Nominal	project String	project_name String	title String	type Nominal
25917	283308.0	A tab for the jmx console that grabs information fro...	Major	FELIX	Felix	A graphical tab to remotely get memory activity	New Feature
25916	283307.0	This MBean / Graphical couple is able to remotely in...	Major	FELIX	Felix	A graphical tab to remotely interact with obr	New Feature
13068	135171.0	We need a .gitignore file under nutch/ so git does no...	Trivial	NUTCH	Nutch	A .gitignore file for Nutch	Improvement
4289	60889.0	Should be work with JSP and Facelets	Major		MyFaces Tobago	A Bean-Tag or CRUD-Tag to create dynamic pages with low effort	New Feature
47466	501275.0	try examination example with exam set 2: after initia...	Major	PLANNER	OptaPlanner	A BestSolutionChangedEvent should be fired when a StartingSolutionInit...	Bug
28177	308725.0	Several people have asked if ManifoldCF supports C...	Major	CONNECTO...	ManifoldCF	A CMIS connector would be helpful	New Feature
6851	85882.0	This bug is a little symetric of SMXCOMP-446 (a CXF...	Major		ServiceMix Components	A CXF-BC provider used with WS-RM sends the CreateSequence request ...	Bug
6721	84487.0	Clob cannot be directly be consumed by Solr. So Jdb...	Minor	SOLR	Solr	A Clobtransformer to read strings from Clob	Improvement
6353	80795.0	This means you cant try and create the collection ag...	Major	SOLR	Solr	A Collection that is only partially created and then deleted will leave pre ...	Bug
39189	416166.0	There should be a standard data converter for java.util.C...	Major	XSTR	XStream	A Currency converter	Improvement
1246	32119.0	A DIV tag that utilized Ajax to obtain its contents. S...	Major		Struts 2	A DIV tag that utilized Ajax to obtain its contents	New Feature
19412	197278.0	See the following comment for detailed description.	Major		Hadoop HDFS	A Distributed and Cooperative NameNode Cluster for a Highly-Available ...	New Feature
1820	38999.0	A few of the projects were missing site.xml site desc...	Minor	WICKET	Wicket	A Few Wicket Projects are Missing site.xml and wicket-site-skin	Bug
51153	556158.0	I have attached a patch that provides a Gradle build ...	Major	GROOVY	groovy	A Gradle build for Groovy	New Feature
52532	575089.0	See JBSEAM-4276.Patch proposed in JBSEAM-4276 i...	Blocker		JBoss Enterprise Appli...	A JBDS generated WAR project with reverse-engineered entities doesnt ...	Bug
58047	645660.0	A patch proposed in JBSEAM-4276 is not applied int...	Blocker		Developer Studio (JBo...	A JBDS generated WAR project with reverse-engineered entities doesnt ...	Bug
38880	413151.0	A JobExecution does not need a start time until it act...	Minor		Spring Batch	A JobExecution does not have a start time until it actually starts	Bug
49068	523929.0	Non-leaf nodes shouldnt have a layerinfo component...	Major	MAPFACES	MapFaces	A LayerInfo component is added even if the node is not a leaf	Bug
7074	88685.0	It would be useful if service-mix provided a store im...	Major		ServiceMix	A MongoDB Store would be usefull	New Feature
6404	81524.0	Id like to fix this for 4 - its a simple fix.	Blocker	SOLR	Solr	A PeerSync attempt to its replicas by a candidate leader should not fail o...	Bug
5323	72970.0	We could do a reference implementation of the Load...	Major	STRATOS	Stratos	A Reference Load Balancer Extension API Implementation for HAProxy	Sub-task
40133	425190.0	A Registry of projects should be maintained. Curren...	Major	UDIG	uDIG	A Registry of projects should be maintained.	Bug
6730	84550.0	This is same as SOLR-139. A new issue is opened so...	Major	SOLR	Solr	A RequestProcessor to support updates	New Feature
35550	376598.0	We need to resolve this issue. Major_case	Major	ABDERA	Abdera	A Runtime Exception Problem	Bug
33218	359243.0	A SOAP 1.2 Message with a SWARef attachment is d...	Major	AXIOM	Axiom	A SOAP 1.2 SWARef Message is denied	Bug
42515	443376.0	see attached screenshot. There should be a cache so...	Major	SONAR	SonarQube	A SQL Select Query is done on Rules table each time a new violation is s...	Improvement
55096	608707.0	The simple plugin tutorial has not been updated fro...	Minor	GEOS	GeoServer	A Simple PlugIn tutorial is not available in the new docs for geoserver ve...	Improvement
6501	82590.0	This is starting to really bug me because tests almos...	Minor	SOLR	Solr	A SolrIndexSearcher can be left open if the executor rejects a task.	Bug
36588	380099.0	It could be useful if a state can extend more than on...	Minor		Spring Web Flow	A State should be able to extend more than one states	Improvement
32198	347338.0	"The web application we are building in this tutorial ...	Trivial	BEEHIVE	Beehive	A Typo in tutorial_pageflow.html	Bug
19000	195423.0	A UDF that can export data to JDBC databases.	Minor	HIVE	Hive	A UDF that can export data to JDBC databases	New Feature
59731	668913.0	Because of new class hierarchy and co-variant return...	Blocker	ISPN	Infinispan	A Weld upgrade is needed	Bug
4290	60890.0	Should be work with JSP and Facelets.	Major		MyFaces Tobago	A Wizard-Tag to create dynamic wizard pages with low effort	New Feature
2059	41215.0	It propably would make sense to have a Set of "Attri...	Minor	VYSPER	VYSPER	A XMLElement might have duplicate attributes which isnt allowed in XML	Bug

ckjm — Chidamber and Kemerer Java Metrics

www.spinellis.gr/sw/ckjm/

(bug_count / project_count) * (blocker_count + critical_count)

QI > 5 = 1 = risky
QI < 5 = 0 = safe

Relation: defect-metrics

No. Numeric	name String	project_count Numeric	bug_count Numeric	new_feature_count Numeric	improvement_count Numeric	sub_task_count Numeric	blocker_count Numeric	critical_count Numeric	major_count Numeric	minor_count Numeric	trivial_count Numeric	NOC Numeric	WMC Numeric	CBO Numeric	RFC Numeric	LCOM Numeric	Ca Numeric	Ce Numeric	ISTBL Numeric	LCOM3 Numeric	LOC Numeric	QI Numeric	Quality Nominal
1	CXF	389.0	167.0	26.0	116.0	20.0	0.0	4.0	312.0	65.0	8.0	173.	3.93	3.45	10.6	24.04	1.69	1.93	0.53	0.51	75.81	1.72	0
2	HBASE	1126.0	471.0	47.0	289.0	132.0	50.0	66.0	699.0	217.0	94.0	402..	3.95	5.13	12.22	37.91	2.58	2.82	0.52	0.49	105...	48.52	1
3	KARAF	316.0	100.0	25.0	88.0	15.0	6.0	5.0	271.0	29.0	5.0	328.0	2.84	1.74	7.68	14.57	0.96	0.96	0.5	0.7	78.18	3.48	0
4	SONAR	577.0	124.0	69.0	304.0	22.0	1.0	17.0	417.0	126.0	16.0	194.0	4.27	2.01	10.3	94.42	1.08	1.08	0.5	0.47	84.06	3.87	0
5	JBPM	447.0	117.0	0.0	0.0	17.0	5.0	15.0	366.0	46.0	1.0	178.0	2.92	4.48	12.61	8.19	0.99	3.51	0.78	0.82	121...	5.23	1
6	FABRICTHREE	209.0	70.0	87.0	47.0	1.0	7.0	1.0	160.0	29.0	2.0	544.0	2.47	4.48	9.25	5.57	0.82	3.75	0.82	0.51	61.59	6.03	1
7	NUTCH	250.0	67.0	30.0	96.0	35.0	1.0	5.0	128.0	87.0	29.0	534.0	3.17	5.42	13.41	7.63	1.48	4.06	0.73	0.48	103...	1.61	0
8	CMIS	121.0	41.0	14.0	39.0	9.0	1.0	4.0	79.0	36.0	1.0	908.0	4.61	3.35	10.51	106.48	1.15	2.29	0.67	0.54	85.06	1.69	0
9	PIG	336.0	148.0	40.0	89.0	28.0	8.0	11.0	265.0	38.0	14.0	466.0	2.59	3.73	10.46	7.03	0.42	3.35	0.89	0.72	75.27	8.37	1
10	OPENEJB	291.0	73.0	29.0	43.0	93.0	2.0	1.0	258.0	28.0	2.0	166...	3.34	3.78	11.33	18.68	1.65	2.24	0.58	0.6	92.42	0.75	0
11	OPENNLP	233.0	33.0	17.0	123.0	3.0	0.0	0.0	117.0	108.0	8.0	744.0	2.83	4.43	9.51	8.83	2.1	2.4	0.53	0.57	77.17	0.00	0
12	JBRULES	376.0	127.0	0.0	0.0	12.0	5.0	16.0	325.0	25.0	2.0	387...	6.1	6.72	12.95	80.91	2.91	4.02	0.58	0.58	92.85	7.09	1
13	ARQ	215.0	40.0	0.0	0.0	3.0	2.0	0.0	171.0	27.0	1.0	670...	3.9	6.04	9.25	40.67	3.13	3.18	0.5	0.59	73.99	2.23	0
14	MILYN	127.0	30.0	35.0	27.0	10.0	3.0	2.0	106.0	15.0	1.0	154...	3.16	4.34	9.41	13.5	2.07	2.37	0.53	0.59	65.12	1.18	0
15	COCOON	101.0	56.0	6.0	22.0	6.0	1.0	3.0	73.0	22.0	2.0	234.0	3.88	3.68	12.84	19.43	0.65	3.18	0.83	0.46	116...	2.22	0
16	AVRO	206.0	65.0	45.0	77.0	5.0	7.0	2.0	158.0	34.0	6.0	358.0	3.61	3.34	9.39	16.23	1.74	1.86	0.52	0.59	76.92	2.84	0
17	BOOKKEEPER	209.0	82.0	14.0	57.0	45.0	11.0	3.0	153.0	36.0	6.0	272...	4.78	4.51	9.95	59.14	2.38	2.63	0.53	0.48	67.93	5.49	1
18	MAHOUT	155.0	52.0	16.0	65.0	6.0	0.0	0.0	88.0	52.0	15.0	176...	3.14	4.79	10.8	8.17	1.56	3.36	0.68	0.47	75.14	0.00	0
19	SHINDIG	214.0	88.0	25.0	79.0	6.0	3.0	2.0	146.0	54.0	9.0	448.0	2.73	2.77	8.07	8.81	1.04	1.89	0.65	0.6	49.68	2.06	0
20	SQOOP	228.0	66.0	30.0	53.0	45.0	0.0	1.0	171.0	26.0	17.0	950.0	2.79	3.7	9.95	36.08	1.27	2.49	0.66	0.68	75.5	4.05	0
21	ISPN	414.0	126.0	0.0	0.0	10.0	21.0	23.0	307.0	54.0	6.0	238...	4.72	6.01	11.42	73.24	3.0	3.4	0.53	0.55	72.94	13.39	1
22	QPID	734.0	326.0	65.0	242.0	31.0	16.0	9.0	530.0	159.0	20.0	544.0	5.05	4.69	13.35	77.9	1.68	3.33	0.66	0.57	96.75	11.11	1
23	SWITCHYARD	390.0	104.0	0.0	0.0	17.0	18.0	15.0	314.0	35.0	5.0	168.0	3.19	1.71	8.08	16.99	0.85	0.93	0.52	0.58	54.17	8.8	1
24	CAY	116.0	31.0	0.0	47.0	1.0	1.0	1.0	77.0	25.0	12.0	273...	6.03	5.55	10.59	69.17	2.88	3.0	0.51	0.56	122...	0.53	0
25	TUSCANY	533.0	236.0	65.0	147.0	12.0	21.0	13.0	382.0	102.0	15.0	38.0	5.42	10.63	2.71	0.55	4.97	0.9	0.59	53.37	15.05	1	
26	TEIID	467.0	182.0	0.0	0.0	36.0	12.0	17.0	338.0	81.0	6.0	264.0	3.47	4.78	11.51	24.76	1.27	3.9	0.75	0.66	86.65	11.31	1
27	OFBIZ	832.0	265.0	46.0	396.0	77.0	4.0	8.0	303.0	396.0	121.0	18.0	3.67	1.11	16.83	10.83	0.67	0.67	0.5	0.44	141...	3.82	0
28	SOLR	534.0	168.0	95.0	193.0	22.0	6.0	7.0	262.0	209.0	50.0	215...	3.06	5.34	11.18	14.04	1.92	3.78	0.66	0.51	93.67	4.09	0
29	AMBARI	827.0	479.0	12.0	84.0	129.0	6.0	21.0	771.0	22.0	7.0	52.0	1.88	1.13	2.42	1.75	0.54	0.6	0.53	0.81	7.23	15.64	1
30	CACTUS	221.0	58.0	59.0	44.0	12.0	5.0	10.0	166.0	35.0	5.0	274.0	5.22	5.31	14.98	50.3	1.56	3.83	0.71	0.6	138...	3.94	0

- **WMC:** Weighted methods per class
- **DIT:** Depth of Inheritance Tree
- **NOC:** Number of Children
- **CBO:** Coupling between object classes
- **RFC:** Response for a Class
- **LCOM:** Lack of cohesion in methods

In addition it also calculates for each class

- **Ca:** Afferent couplings
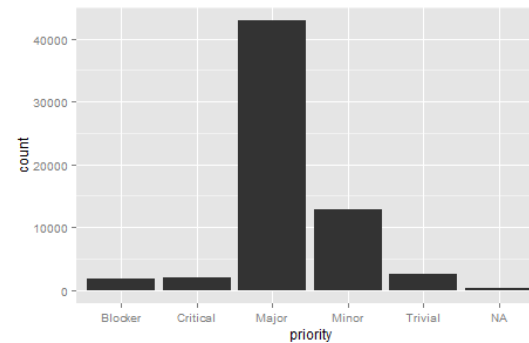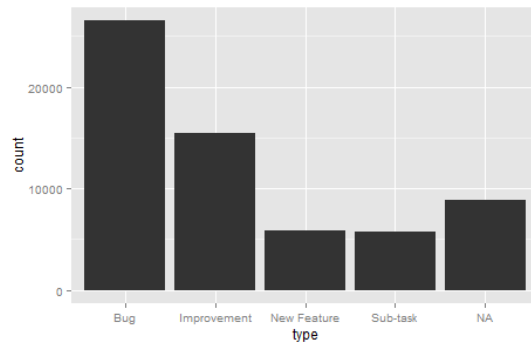- **NPM:** Number of public methods

Component Prediction

More ▼

Both Alex and Steve emphasized that research on component prediction seems worthwhile from the industry viewpoint.

Predictor-input: report_description + report_title
Predictor-output: report_project-name

Subclass Discovery

More ▼

Predicting Bugs
- semi-supervised. use the given types to train the model. then predict the missing labels [bug or new feature]

Predictor-input: report_description + report_title + report_priority [not sure if this can also be added for input]

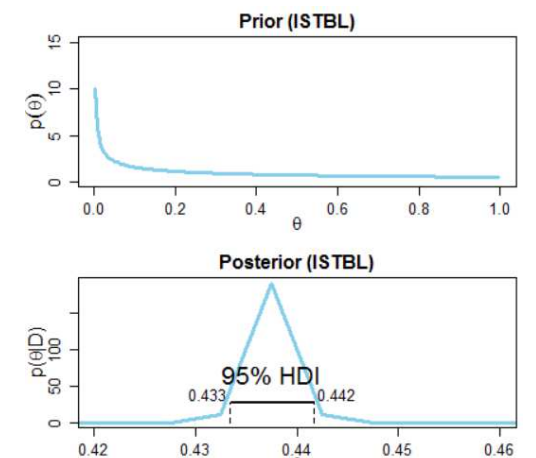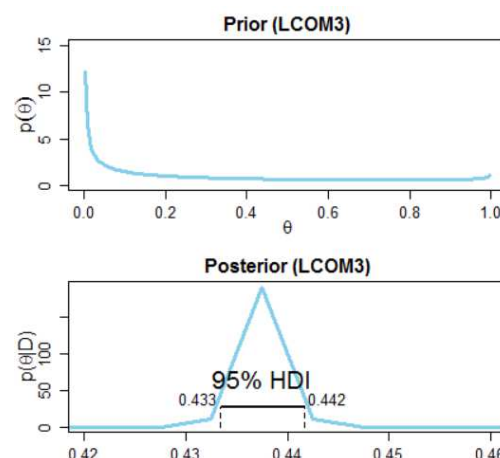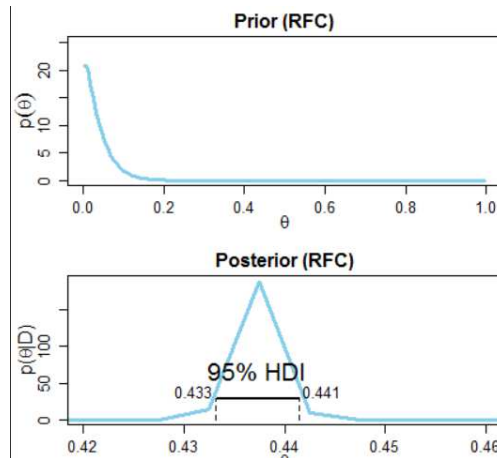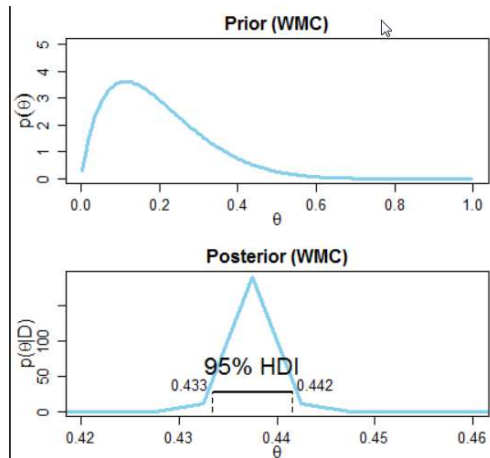Predictor-output-1: report_type [bug or new feature]

Predictor-output-2: report_priority [optional, if not added for input and if time allows to predict as output]

```
# Total  PROJECTS     55642

# Total  BUG          24344

# Total  NOT-BUG      31298

post = BernBeta( c(1.88,7.88), c(rep(1,24344),rep(0,31298)))
```
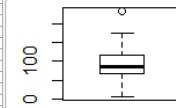
Relation: defect-metrics

No.	name	project_count	bug_count	new_feature_count	improvement_count	sub_task_count	blocker_count	critical_count	major_count	minor_count	trivial_count	NOC	WMC	CBO	RFC	LCOM	Ca	Ce	ISTBL	LCOM3	LOC	QI	Quality
	String	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Nominal
1	CXF	389.0	167.0	26.0	116.0	20.0	0.0	4.0	312.0	65.0	8.0	173...	3.93	3.45	10.6	24.04	1.69	1.93	0.53	0.51	75.81	1.720	
2	HBASE	1126.0	471.0	47.0	289.0	132.0	50.0	66.0	699.0	217.0	94.0	402...	3.95	5.13	12.22	37.91	2.58	2.82	0.52	0.49	105...	48.521	
3	KARAF	316.0	100.0	25.0	88.0	15.0	6.0	5.0	271.0	29.0	5.0	328.0	2.84	1.74	7.68	14.57	0.96	0.96	0.5	0.7	78.18	3.480	
4	SONAR	577.0	124.0	69.0	304.0	22.0	1.0	17.0	417.0	126.0	16.0	194.0	4.27	2.01	10.3	94.42	1.08	1.08	0.5	0.47	84.06	3.870	
5	JBPM	447.0	117.0	0.0	0.0	17.0	5.0	15.0	366.0	46.0	1.0	178.0	2.92	4.48	12.61	8.19	0.99	3.51	0.78	0.82	121...	5.231	
6	FABRICTHREE	209.0	70.0	87.0	47.0	1.0	7.0	11.0	160.0	29.0	2.0	544.0	2.47	4.48	9.25	5.57	0.82	3.75	0.82	0.51	61.59	6.031	
7	NUTCH	250.0	67.0	30.0	96.0	35.0	1.0	5.0	128.0	87.0	29.0	534.0	3.17	5.42	13.41	7.63	1.48	4.06	0.73	0.48	103...	1.610	
8	CMIS	121.0	41.0	14.0	39.0	9.0	1.0	4.0	79.0	36.0	1.0	908.0	4.61	3.35	10.51	106.48	1.15	2.29	0.67	0.54	85.06	1.690	
9	PIG	336.0	148.0	40.0	89.0	28.0	8.0	11.0	265.0	38.0	14.0	466.0	2.59	3.73	10.46	7.03	0.42	3.35	0.89	0.72	75.27	8.371	
10	OPENEJB	291.0	73.0	29.0	43.0	93.0	2.0	1.0	258.0	28.0	2.0	166...	3.34	3.78	11.33	18.68	1.65	2.24	0.58	0.6	92.42	0.750	
11	OPENNLP	233.0	33.0	17.0	123.0	3.0	1.0	0.0	117.0	108.0	8.0	744.0	2.83	4.43	9.51	8.83	2.1	2.4	0.53	0.57	77.17	0.00	
12	JBRULES	376.0	127.0	0.0	0.0	12.0	5.0	16.0	325.0	25.0	2.0	166...	6.1	6.72	12.95	80.91	2.91	4.02	0.58	0.58	92.85	7.091	
13	ARQ	215.0	40.0	0.0	0.0	3.0	2.0	10.0	171.0	27.0	1.0	670...	3.9	6.04	9.25	40.67	3.13	3.18	0.5	0.59	73.99	2.230	
14	MILYN	127.0	30.0	0.0	35.0	27.0	0.0	3.0	106.0	15.0	1.0	154...	3.16	4.34	9.41	13.5	2.07	2.37	0.53	0.59	65.12	1.180	
15	COCOON	101.0	56.0	6.0	22.0	6.0	1.0	3.0	73.0	22.0	2.0	234.0	3.88	3.68	12.84	19.43	0.65	3.18	0.83	0.46	116...	2.220	
16	AVRO	206.0	65.0	45.0	77.0	5.0	7.0	2.0	158.0	34.0	5.0	358.0	3.61	3.34	9.39	16.23	1.74	1.86	0.52	0.59	76.92	2.840	
17	BOOKKEEPER	209.0	82.0	14.0	57.0	45.0	11.0	3.0	153.0	36.0	6.0	272...	4.78	4.51	9.95	59.14	2.38	2.63	0.53	0.48	67.93	5.491	
18	MAHOUT	155.0	52.0	16.0	65.0	6.0	0.0	0.0	88.0	52.0	15.0	176...	3.14	4.79	10.8	8.17	1.56	3.36	0.68	0.47	75.14	0.00	
19	SHINDIG	214.0	88.0	25.0	79.0	6.0	3.0	2.0	146.0	54.0	9.0	448.0	2.73	2.77	8.07	8.81	1.04	1.89	0.65	0.56	49.68	2.060	
20	SQOOP	228.0	66.0	30.0	53.0	45.0	11.0	3.0	171.0	26.0	17.0	950.0	2.79	3.7	9.95	36.08	1.27	2.49	0.66	0.68	75.5	4.050	
21	ISPN	414.0	126.0	0.0	0.0	10.0	21.0	23.0	307.0	54.0	6.0	238...	4.72	6.01	11.42	73.24	3.0	3.4	0.53	0.55	72.94	13.391	
22	QPID	734.0	326.0	65.0	242.0	31.0	16.0	9.0	530.0	159.0	20.0	544.0	5.05	4.69	13.35	77.9	1.68	3.33	0.66	0.57	96.73	11.11	
23	SWITCHYARD	390.0	104.0	0.0	0.0	17.0	18.0	15.0	314.0	35.0	5.0	168.0	3.19	1.71	8.08	16.99	0.85	0.93	0.52	0.58	54.17	8.81	
24	CAY	116.0	31.0	0.0	47.0	1.0	1.0	1.0	77.0	25.0	12.0	273...	6.03	5.55	10.59	69.17	2.88	3.0	0.51	0.56	122...	0.530	
25	TUSCANY	533.0	236.0	65.0	147.0	12.0	21.0	13.0	382.0	102.0	15.0	38.0	1.89	5.42	10.63	2.71	0.55	4.97	0.9	0.59	53.37	15.051	
26	TEIID	467.0	182.0	0.0	0.0	36.0	12.0	17.0	338.0	81.0	6.0	264.0	3.47	4.78	11.51	24.76	1.27	3.9	0.75	0.42	86.65	11.31	
27	OFBIZ	832.0	265.0	46.0	396.0	77.0	4.0	8.0	303.0	396.0	121.0	18.0	3.67	11.16.83	10.83	0.67	0.67	0.5	0.44	141...	3.820	
28	SOLR	534.0	168.0	95.0	193.0	22.0	6.0	7.0	262.0	209.0	50.0	215...	3.06	5.34	11.18	14.04	1.92	3.78	0.66	0.54	93.67	4.090	
29	AMBARI	827.0	479.0	12.0	84.0	129.0	6.0	21.0	771.0	22.0	7.0	52.0	1.88	1.13	2.42	5.77	0.54	0.6	0.53	0.81	7.23	15.641	
30	CACTUS	221.0	58.0	59.0	44.0	12.0	5.0	10.0	166.0	35.0	5.0	274.0	5.22	5.31	14.98	50.3	1.56	3.83	0.71	0.6	138...	3.940	

Prior (WMC) · Prior (RFC) · Prior (LCOM3) · Prior (ISTBL)

Posterior (WMC) — 95% HDI 0.433 – 0.442

Posterior (RFC) — 95% HDI 0.433 – 0.441

Posterior (LCOM3) — 95% HDI 0.433 – 0.442

Posterior (ISTBL) — 95% HDI 0.433 – 0.442

Relation: defect-metrics

| No.	name	project_count	bug_count	new_feature_count	improvement_count	sub_task_count	blocker_count	critical_count	major_count	minor_count	trivial_count	NOC	WMC	CBO	RFC	LCOM	Ca	Ce	ISTBL	LCOM3	LOC	QI	Quality
		Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Nominal
1	CXF	389.0	167.0	26.0	116.0	20.0	0.0	4.0	312.0	65.0	8.0	173...	3.93	3.45	10.6	24.04	1.69	1.93	0.53	0.51	75.81	1.72	0
2	HBASE	1126.0	471.0	47.0	289.0	132.0	50.0	66.0	699.0	217.0	94.0	402...	3.95	5.13	12.22	37.91	2.58	2.82	0.52	0.49	105...	48.52	1
3	KARAF	316.0	100.0	25.0	88.0	15.0	6.0	5.0	271.0	29.0	5.0	328.0	2.84	1.74	7.68	14.57	0.96	0.96	0.5	0.7	78.18	3.48	0
4	SONAR	577.0	124.0	69.0	304.0	22.0	1.0	17.0	417.0	126.0	16.0	194.0	4.27	2.01	10.3	94.42	1.08	1.08	0.5	0.47	84.06	3.87	0
5	JBPM	447.0	117.0	0.0	0.0	17.0	5.0	15.0	366.0	46.0	1.0	178.0	2.92	4.48	12.61	8.19	0.99	3.51	0.78	0.82	121...	5.23	1
6	FABRICTHREE	209.0	70.0	47.0	1.0	7.0	11.0		160.0	29.0	2.0	544.0	2.47	4.48	9.25	5.57	0.82	3.75	0.82	0.51	61.59	6.03	1
7	NUTCH	250.0	67.0	30.0	96.0	35.0	1.0	5.0	128.0	87.0	29.0	534.0	3.17	5.42	13.41	7.63	1.48	4.06	0.73	0.48	103...	1.61	0
8	CMIS	121.0	41.0	14.0	39.0	9.0	1.0	4.0	79.0	36.0	1.0	908.0	4.61	3.35	10.51	106.48	1.15	2.29	0.67	0.54	85.06	1.69	0
9	PIG	336.0	148.0	40.0	89.0	28.0	8.0	11.0	265.0	38.0	14.0	466.0	2.59	3.73	10.46	7.03	0.42	3.35	0.89	0.72	75.27	8.37	1
10	OPENEJB	291.0	73.0	29.0	43.0	93.0	2.0	1.0	258.0	28.0	2.0	166...	3.34	3.78	11.33	18.68	1.65	2.24	0.58	0.6	92.42	0.75	0
11	OPENNLP	233.0	33.0	17.0	123.0	3.0	0.0	0.0	117.0	108.0	8.0	744.0	2.83	4.43	9.51	8.83	2.1	2.4	0.53	0.57	77.17	0.00	0
12	JBRULES	376.0	127.0	0.0	0.0	12.0	5.0	16.0	325.0	25.0	2.0	387...	6.1	6.72	12.95	80.91	2.91	4.02	0.58	0.58	92.85	7.09	1
13	ARQ	215.0	40.0	0.0	0.0	3.0	2.0	10.0	171.0	27.0	1.0	670...	3.9	6.04	9.25	40.67	3.13	3.18	0.5	0.59	73.99	2.23	0
14	MILYN	127.0	30.0	35.0	27.0	10.0	3.0	2.0	106.0	15.0	1.0	154...	3.16	4.34	9.41	13.5	2.07	2.37	0.53	0.59	65.12	1.18	0
15	COCOON	101.0	56.0	6.0	22.0	6.0	1.0	3.0	73.0	22.0	2.0	234.0	3.88	3.68	12.84	19.43	0.65	3.18	0.83	0.46	116...	2.22	0
16	AVRO	206.0	65.0	0.0	77.0	5.0	7.0	2.0	158.0	34.0	5.0	358.0	3.61	3.34	9.39	16.23	1.74	1.86	0.52	0.59	76.92	2.84	0
17	BOOKEEPER	209.0	82.0	14.0	57.0	45.0	11.0	3.0	153.0	36.0	6.0	272...	4.78	4.51	9.95	59.14	2.38	2.63	0.53	0.48	67.93	5.49	1
18	MAHOUT	155.0	52.0	16.0	65.0	6.0	0.0	0.0	88.0	52.0	15.0	176...	3.14	4.79	10.8	8.17	1.56	3.36	0.68	0.47	75.14	0.00	0
19	SHINDIG	214.0	88.0	25.0	79.0	6.0	3.0	2.0	146.0	54.0	9.0	448.0	2.73	2.77	8.07	8.81	1.04	1.89	0.65	0.6	49.68	2.06	0
20	SQOOP	228.0	66.0	30.0	53.0	45.0	11.0	3.0	171.0	26.0	17.0	950.0	2.79	3.7	9.95	36.08	1.27	2.49	0.66	0.68	75.5	4.05	0
21	ISPN	414.0	126.0	0.0	0.0	10.0	21.0	23.0	307.0	54.0	6.0	238...	4.72	6.01	11.42	73.24	3.0	3.4	0.53	0.55	72.94	13.39	1
22	QPID	734.0	326.0	65.0	242.0	31.0	16.0	9.0	530.0	159.0	20.0	544.0	5.05	4.69	13.35	77.9	1.68	3.33	0.66	0.57	96.75	11.1	1
23	SWITCHYARD	390.0	104.0	0.0	17.0	18.0	15.0		314.0	35.0	5.0	168.0	3.19	1.71	8.08	16.99	0.85	0.93	0.52	0.59	53.37	8.8	1
24	CAY	116.0	31.0	0.0	47.0	1.0	1.0	1.0	77.0	25.0	12.0	273...	6.03	5.55	10.59	69.17	2.88	3.0	0.51	0.56	122...	0.53	0
25	TUSCANY	533.0	236.0	65.0	147.0	12.0	21.0	13.0	382.0	102.0	15.0	38.0	1.89	5.42	10.63	2.71	0.55	4.97	0.9	0.59	53.37	15.05	1
26	TEIID	467.0	182.0	0.0	0.0	36.0	12.0	17.0	338.0	81.0	6.0	264.0	3.47	4.78	11.51	24.76	1.27	3.9	0.75	0.42	86.65	11.3	1
27	OFBIZ	832.0	265.0	46.0	396.0	77.0	4.0	8.0	303.0	396.0	121.0	18.0	3.67	1.11	16.83	10.83	0.67	0.67	0.5	0.44	141...	3.82	0
28	SOLR	534.0	168.0	95.0	193.0	22.0	6.0	7.0	262.0	209.0	50.0	215...	3.06	5.34	11.18	14.04	1.92	3.78	0.66	0.54	93.67	4.09	0
29	AMBARI	827.0	479.0	12.0	84.0	129.0	6.0	21.0	771.0	22.0	7.0	52.0	1.88	1.13	2.42	5.77	0.54	0.6	0.53	0.81	7.23	15.64	1
30	CACTUS	221.0	58.0	59.0	44.0	12.0	5.0	10.0	166.0	35.0	5.0	274.0	5.22	5.31	14.98	50.3	1.56	3.83	0.71	0.6	138...	3.94	0
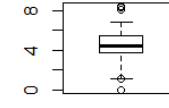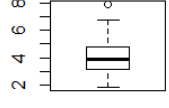
LOC   ISTBL   CBO   WMC   WMC
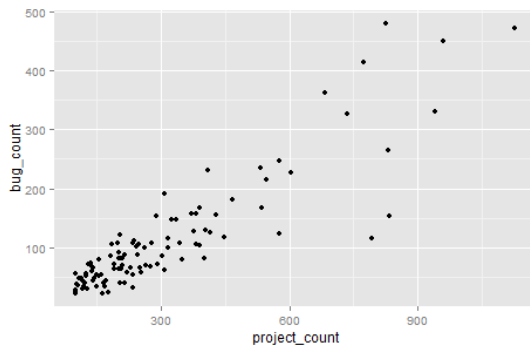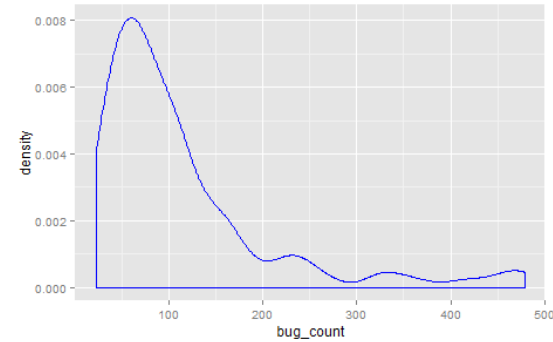

No of bugs increases with project size
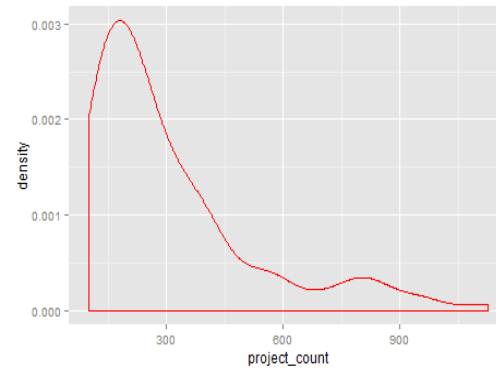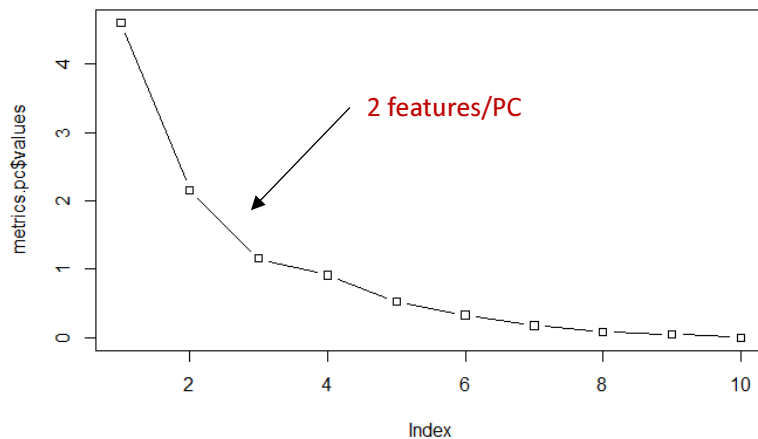

Both pdfs are skewed to the right. Average issues reported per project are on a range of 150-250.

Scree plot of software metrics principal components

2 features/PC

Biplot of software metrics principal components with orthogonal roation


```
> metrics.cor
        NOC   WMC   CBO   RFC  LCOM    Ca    Ce ISTBL LCOM3   LOC
NOC    1.00  0.35  0.49  0.10  0.49  0.68  0.20 -0.36 -0.17  0.18
WMC    0.35  1.00  0.48  0.65  0.77  0.46  0.32 -0.11 -0.21  0.64
CBO    0.49  0.48  1.00  0.64  0.39  0.68  0.87  0.00 -0.37  0.57
RFC    0.10  0.65  0.64  1.00  0.38  0.24  0.69  0.17 -0.35  0.90
LCOM   0.49  0.77  0.39  0.38  1.00  0.53  0.16 -0.25 -0.11  0.46
Ca     0.68  0.46  0.68  0.24  0.53  1.00  0.23 -0.65 -0.35  0.31
Ce     0.20  0.32  0.87  0.69  0.16  0.23  1.00  0.42 -0.29  0.55
ISTBL -0.36 -0.11  0.00  0.17 -0.25 -0.65  0.42  1.00  0.17  0.01
LCOM3 -0.17 -0.21 -0.37 -0.35 -0.11 -0.35 -0.29  0.17  1.00 -0.28
LOC    0.18  0.64  0.57  0.90  0.46  0.31  0.55  0.01 -0.28  1.00
```

```
> metrics.pc <- principal(metrics, nfactors=2, rotate="varimax")
```

```
Principal Components Analysis
Call: principal(r = metrics, nfactors = 2, rotate = "varimax")
Standardized loadings (pattern matrix) based upon correlation matrix

        RC1   RC2   h2   u2

RFC    0.93  0.04 0.86 0.14
Ce     0.88 -0.11 0.79 0.21
LOC    0.83  0.17 0.72 0.28
CBO    0.80  0.36 0.77 0.23
WMC    0.63  0.47 0.62 0.38

LCOM   0.40  0.64 0.57 0.43
LCOM3 -0.37 -0.26 0.20 0.80
ISTBL  0.34 -0.81 0.77 0.23
Ca     0.27  0.89 0.86 0.14
NOC    0.18  0.75 0.59 0.41

                       RC1   RC2
SS loadings           3.88  2.88
Proportion Var        0.39  0.29
```

Factor plot for software metrics principle components

Relation: defect-metrics


```
> model <- glm(fmla, data=train, family=binomial(link="logit"))
```

```
> cM <- table(truth=test$Quality, prediction=test$pred>0.1)
```

```
> print(cM)
        prediction
truth FALSE TRUE
    0     1   10
    1     0   14
```

Zero type 2 error

```
> smp_size <- floor(0.60 * nrow(metrics))
```

25	62

Conclusion: metrics regression performance is 70% correct. with zero false negative or type 2 error at regression threshold 0.1.

Continue data minning and model comparison with

- DT, Näive Bayes, SVM (?), ensamble techniques

Summarize the findings and publish literature survey

- Collect publication requirements
- Preliminary selected IST for publication target

Information and Software Technology
Supports Open Access | About this Journal | Sample Issue
Formerly known as **Data Processing;**
Get new article feed
Get new Open Access article feed
Subscribe to new article alerts
Add to Favorites
Copyright © 2015 Elsevier B.V. All rights reserved

- Project roadmap revesit

	Q1 2015	Q2 2015	Q3 2015	Q4 2015	Q1 2016	Q2 2016	Q3 2016	Q4 2016	Q1 2017	Q2 2017	Q3 2017	Q4 2017	Q1 2018	Q2 2018	Q3 2018	Q4 2018
Phase 1: 960 hrs (aprox.)																
Theoretical Background	WIP	TODO	TODO													
Bayesian Data Analysis	WIP															
Infer Network Structure with K2	TODO															
Phase 2: 2400 hrs (aprox.)																
Component Prediction				TODO	TODO	TODO										
Subclass Discovery					TODO	TODO										
Phase 3: 2400 hrs (aprox.)																
Active Learning							TODO	TODO	TODO							
Thesis Writing			TODO		TODO		TODO	TODO	TODO	TODO	TODO	TODO				
Color Code	TODO	WIP	DONE													

- Project plan submitted
- Following Supervisor Meeting Reminder received from

From: admin-postmaster@brunel.ac.uk
Sent: 15 March 2015 01:16:25 (UTC+02:00) Helsinki, Kyiv, Riga, Sofia, Tallinn, Vilnius
To: Mohammad Sharif Hanif
Subject: Supervisor Meeting Reminder

Dear Mohammad Sharif Hanif,

This email is a reminder that your supervisor meeting on 12/Mar/2015 is now overdue. You should upload your meeting documentation via eVision to ensure that your supervisor can review it.

PLEASE DO NOT REPLY TO THIS EMAIL AS THIS IS AN UNMONITORED ACCOUNT.

My Supervisor Meetings 2014/5		
Meeting Number	Expected Meeting Date	Meeting Status
1	12/Mar/2015	Supervisor Sign-Off Required

- In 2-3 weeks

Result Expected:
- Predictive permormance model comparison
- First publication of lit-survey requirements collection started