

Contents

1	Abstract	6
2	Introduction	8
	2.1 Research Goals	11
3	Related Work	12
4	Methodology and Results	15
	4.1 Time series analysis of open source developers collab- oration communication sentiments	16
	4.2 Social network analysis of open source developers in- teraction graphs	30
5	Conclusion	60
6	Threats to Validity	61
Appendices		62
1	Appendix A: List of all projects forked in socially-related Un- desirable Forking (U.F.) Category	63
2	Appendix B: Initial study: Temporal analysis using the network- specific measurement approach	65
	2.1 Visualization	67
	2.2 Initial study results and discussion	67
3	Appendix C: Mathematical Definition of Effects in the Statis- tical Model	74
	3.1 Structural Effects for the Objective Function	74
	3.2 Behavior-related Effects	76

List of Figures

1	Time series with outliers detected for projects (From top) Kamailio, ffmpeg, Amarok. Our analysis did not find any outliers for Apache CouchDB.	18
2	Time series with outliers detected for projects (From top) Pidgin, MPlayer, Asterisk, and freeglut. Our analysis did not succussfully analyse the rdesktop data.	19
3	Time series with outliers detected for projects (From top) Ceph, OpenStack Neutron, and GlusterFS.	20
4	Time Series of sentiments in developers' messages for projects (From top to bottom, left to right) Kamailio, ffmpeg, Amarok, Apache CouchDB, Pidgin, MPlayer.	23
5	Time Series of sentiments in developers' messages for projects (From top to bottom, left to right) Asterisk, rdesktop, freeglut, Ceph, OpenStack Neutron, and GlusterFS	24
6	Time series decomposed into trend, seasonality and residual time Series for projects (From top to bottom, left to right) Kamailio, ffmpeg, Amarok, Apache CouchDB.	25
7	Time series decomposed into trend, seasonality and residual time Series for projects (From top to bottom, left to right) Pidgin, MPlayer, Asterisk, rdesktop.	26
8	Time series decomposed into trend, seasonality and residual time Series for projects (From top to bottom, left to right) freeglut, Ceph, OpenStack Neutron, and GlusterFS	27
9	The methodology in a glance	31
10	The methodology overview	32
11	Heat-map color-coded examples Degree, Closeness, Betweenness, Eigenvector centralities	66

12	The number of nodes and edges over time, as network-specific measurements	68
13	The network diameter change over time, as network-specific measurements	69
14	Snapshots from video visualization of Kamilio project's graph	70
15	Kamilio top contributors' betweenness centralities and network diameter over time	70
16	Amarok project's top contributors' betweenness centralities and network diameter over time	71
17	Asterisk project's top contributors' betweenness centralities and network diameter over time	72

List of Tables

1	The main reasons for forking	10
2	The frequency of main reasons for forking	10
3	The socially-related forking reasons	11
4	The behavioral measures used by Asur et al. [1]	14
5	The measures of diversity used by Kunegis et al. [30]	15
6	Anomalies detected in the time series of sentiments for project Kamailio	22
7	Anomalies detected in the time series of sentiments for project ffmpeg	22
8	Anomalies detected in the time series of sentiments for project Amarok	22
9	Anomalies detected in the time series of sentiments for project Pidgin	28
10	Anomalies detected in the time series of sentiments for project MPlayer	28
11	Anomalies detected in the time series of sentiments for project Asterisk	28
12	Anomalies detected in the time series of sentiments for project rdesktop	28
13	Anomalies detected in the time series of sentiments for project freeglut	29
14	List of projects in U.F. H.F., and No.F. categories selected based on the criteria described in section 4.2.1 for our study .	34
15	Estimates parameters in the model for Project Kamailio . . .	49
16	Estimates parameters in the model for Project ffmpeg	50
17	Estimates parameters in the model for Project Amarok	51
18	Estimates parameters in the model for Project Apache CouchDB	52
19	Estimates parameters in the model for Project Pidgin	53

20	Estimates parameters in the model for Project MPlayer	54
21	Estimates parameters in the model for Project rdesktop	55
22	Estimates parameters in the model for Project freeglut	56
23	Estimates parameters in the model for Project Ceph	57
24	Estimates parameters in the model for Project OpenStack Neutron	58
25	Estimates parameters in the model for Project GlusterFS . . .	59
26	List of all projects forked because of “ <i>personal differences among the developer team</i> ” (U.F.)	63
27	List of all projects forked because of the need for “ <i>more community- driven development</i> ” (U.F.)	64