



BullyBlocker

Detecting Cyberbullying Victimization Risk Through an Interdisciplinary Identification Model

Ashley C. L. Trow, Linle Jiang, Carmen Sanchez, Victoria Delgadillo, Deborah Hall, Yasin Silva

Arizona State University

BullyBlocker App

This project is an interdisciplinary collaboration drawing on empirical psychological findings and emerging techniques in computer science to combat cyberbullying among teens. A primary objective is to develop an automated identification model that provides parents with an estimate of the likelihood that their teen is being cyberbullied through social media. The identification model, which is incorporated into a mobile app called BullyBlocker, uses data extracted from a teen's social media profile and comments or posts from members of their social network, as well as information provided by parents, to calculate an estimated likelihood that the teen is, or may be, a victim of cyberbullying.

The newest version of the BullyBlocker app is currently available for parents to download for free from the Apple Store. To use the app, parents must enter the login information for their teen's Facebook account. Once a "user profile" is created for each teen that the parent is interested in monitoring for cyberbullying risk, the app extracts data from the teen's Facebook profile, including comments made by other users on the teen's wall posts, photos, and videos. Based on the extracted data and information provided by the parent, the app calculates a Bullying Rank – a score ranging from o to 100 estimating the likelihood that the teen is being cyberbullied - based on the underlying identification model described herein. Once a Bullying Rank is calculated, the parent is provided with a customized list of resources – ranging from crisis hotlines to advocacy websites and online educational materials – that are selected based on the information detected by the identification model.

	arrier 🗢 11:04 AM 📼	Carrier 🗢 10:39 AM		Carrier 🗢 10:47 AM		Carrier 🗢	10:54 AM
PROFLE NFORMATION Ethicity Other > Race Black or African American > Mental health history > Past history of being bullied Monetto mentals 0 Race Insulting Feed Content: 0 Insulting Photo comments: 2 Insulting Photo comments: 0 Name: Karia Winther Bulying Rank:: Low Risk O 6 Warning Signs: Insulting Photo comments: 2 Insulting Photo comments: 0 New Neighborhood Factor: 100 New School Factor: 100 New Reighborhood Factor: 100 New School Factor: 100 Secuel Orientalito Factor: 100 New Reighborhood Factor: <t< td=""><td>C Facebook Accounts</td><td>Facebook Accounts</td><td></td><td>Facebook Accounts</td><td></td><td>Faceboo</td><td>ok Accounts</td></t<>	C Facebook Accounts	Facebook Accounts		Facebook Accounts		Faceboo	ok Accounts
PPORTALE INFORMATION Ethicity Other > Race Black or African American > Mental health history > Past history of being bullied like, not to my knowsdrege > Insulting Feed Content: 0 Bialty internet usage Less than 1 hour > Winner Bizgling Factor: 100 New Neighborhood Factor: 100 Sexual Orientation Factor: 100 Number of Days Sc				4		Recomm	ended Resources
Ethnicity Oter > Race Black or African American > Mental health history > Past history of being bullied Ne, not to my knowledge > Past history of being bullied Ne, not to my knowledge > Daily internet usage Less than 1 hour > New Neighborhood Factor: 100 New Neighborhood Factor: 000 Gender Factor: 100 Number of Days Scanneel: 90	PROFILE INFORMATION	Name: Maria Winther		Name: Constance Winther		(AR)	Drug Rebab
Currently Outer 3 Race Black or African American 3 Mental health history > Past history of being bullied Me, not termy knowledge 2 Past disciplinary issues > Daily internet usage Less than 1 hour 3 Warning Signs: Insulting Fedd Content: 0 Insulting Photo comments: 2 Insulting Photo comments: 0 Vulnerability Factors: New Neighborhood Factor: 1.00 New School Factor: 0.00 Gender Factor: 0.00 Bullying History Factor: 0.00 Bullying History Factor: 0.00 Gender Factor: 0.00 Gender Factor: 0.00 Bullying History Factor: 0.00 Bullying History Factor: 0.00 Gender Factor: 0.00 Gender Factor: 0.00 Number of Days Scanned: 90	Ethnicity Other	Bullving Bank: Low Risk		Bullving Bank: High Bisk			Didg Kellab
Race Black or African American > 6 100 6 100 83 100	Curier 2	Dullying Kalik. Low Kisk	100		100		Help Your Teen Now
Warning Signs: Insulting Feed Content: 0 Vulnerability Factors: 0 New Neighborhood Factor: 1.00 New School Factor: 1.00 Bed Juling History Factor: 0.00 Mental Health History Factor: 0.00 Bullying History Factor: 0.00 Bercel: 0 Number of Days Scanned: 90 Number	Race Black or African American >	6	100	83	100		
iental health history insulting Golds. iast history of being bullied Ne, not omy knowledge > ast history of being bullied Ne, not omy knowledge > iast history issues > iast history issues > iast history issues > ialigi internet usage Less than 1 hour > New Neighborhood Factor: 1.00 New Neighborhood Factor: 0.00 Gender Factor: 0.00 Gender Factor: 0.00 Bullying History Factor: 0.00 Bullying History Factor: 0.00 Bullying History Factor: 0.00 Gender Factor: 0.00 Mental Health History Factor: 0.00 Bullying History Factor: 0.00 Sexual Orientation Factor: 0.00 Mental Health History Factor: 0.00 Sexual Orientation Factor: 0.00 Number of Insults per day:		Warning Signs:		Warning Signs:			Drug Abuse
ast history of being bullied hearts on y knowledge > last disciplinary issues Insulting Pede Content: 0 Insulting Pede Content: 15 More Resources haily internet usage Less than 1 hour > Insulting Video comments: 0 Insulting Pede Content: 15 Insulting Video comments: 35 haily internet usage Less than 1 hour > Insulting Video comments: 0 Insulting Pede Content: 10 Nore Resources New Neighborhood Factor: 1.00 Age Factor: 0.00 Gender Factor: 1.00 Gender Factor: 0.00 Gender Factor: 0.00 Gender Factor: 1.00 Dulying History Factor: 0.00 Gender Factor: 0.00 Gender Factor: 1.00 Bullying History Factor: 0.00 Gender Factor: 0.00 More Resources Born This Way Foundation Past Disciplinary Issues Factor: 0.00 Mental Health History Factor: 0.00 Sexual Orientation Factor: 1.00 Sexual Orientation Factor: 0.00 Number of Days Scanned: 90 Number of Insults per day: 0.02 Number of Insults per day: 0.02 More Fue More Mactor More Mactor <td>Mental health history ></td> <td></td> <td></td> <td>have the stand of the stand</td> <td>45</td> <td>A</td> <td>Revend Bullving</td>	Mental health history >			have the stand of the stand	45	A	Revend Bullving
Part disciplinary issues > Past disciplinary issues > Daily internet usage Less than 1 hour > New Neighborhood Factor: 1.00 New School Factor: 1.00 Age Factor: 0.00 Gender Factor: 0.00 Bullying History Factor: 0.00 Mental Health History Factor: 0.00 Bullying History Factor: 0.00 Mental Health History Factor: 0.00 Insulting Video comments: 2 Number of Days Scanned: 90 Number of Insults per day: 0.02 Potential Bullies: 2 Number of Insults per day: 0.02 Mumber of Insults per day: 0.02 Number of Insults per day:		Insulting Feed Content:	0	Insulting Feed Content:	15		Beyond Bullying
Past disciplinary issues Daily internet usage Less than 1 hour > Insulting Video comments: 0 Vulnerability Factors: New Neighborhood Factor: 1.00 Age Factor: 0.00 Gender Factor: 0.00 Sexual Orientation Factor: 0.00 </td <td>Past history of being balled No, hor to my knowledge</td> <td>Insulting Photo comments:</td> <td>2</td> <td>Insulting Photo comments:</td> <td>35</td> <td>More Res</td> <td>sources</td>	Past history of being balled No, hor to my knowledge	Insulting Photo comments:	2	Insulting Photo comments:	35	More Res	sources
Daily internet usage Less than 1 hour > Vulnerability Factors: New Neighborhood Factor: 1.00 New Neighborhood Factor: 1.00 New Neighborhood Factor: 0.00 Gender Factor: 0.00 Gender Factor: 0.00 Gender Factor: 0.00 Hental Health History Factor: 0.00 Gender Factor: 0.00 Hental Health History Factor: 0.00 New Neighborhood Factor: 1.00 New Neighborhood Factor: 0.00 Gender Factor: 0.00 Hental Health History Factor: 0.00 New Neighborhood Factor: 0.00 Past Disciplinary Issues Factor: 0.00 Internet Usage Factor: 0.25 Race/Ethnicity Factor 0.00 Sexual Orientation Factor: 0.00 Potential Bullies: 2 Number of Insults per day: 0.02 Number of Insults per day: 0.02 Number of Insults per day: 0.80 Number of Insults per day: 0.80 Number of Insults per day: 0.80 Number of Insults per	Past disciplinary issues	Insulting Video comments:	0	Insulting Video comments:	22		The Trevor Project
Value finder led usage Less than 11001 V New Neighborhood Factor: 1.00 New School Factor: 0.00 Gender Factor: 0.00 Gender Factor: 0.00 Method Health History Factor: 0.00 Bullying History Factor: 0.00 Method Health History Factor: 0.00 Bullying History Factor: 0.00 Method Health History Factor: 0.00 New Neighborhood Factor: 1.00 Bullying History Factor: 0.00 New Neighborhood Factor: 0.00 Method Health History Factor: 0.00 Potential Bullies: 2 Number of Days Scanned: 90 Number of Days Scanned: 90 Number of Insults per day: 0.02 New Fieldee New Neighborhood Factor: 0.00 Potential Bullies: 2 Number of Days Scanned: 90 Number of Ins	Daily internet usage	Vulnerability Factors:		Vulnerability Factors:		7	
New School Factor: 1.00 Age Factor: 0.00 Gender Factor: 0.00 Bullying History Factor: 0.00 Bullying History Factor: 0.00 Mem 24 Disciplinary Issues Factor: 0.00 Mem 25 Actor 0.00 Bullying History Factor: 0.00 Mem 24 Disciplinary Issues Factor: 0.00 Past Disciplinary Issues Factor: 0.00 Number of Days Scanned: 90 Number of Insults per day: 0.02 Mumber of Insults per day: 0.02 Mem 26 Factor: 0.02 Mem 26 Factor: 0.02 Mem 26 Factor: 0.02 Mumber of Insults per day: 0.02 Mem 26 Factor: 0.02 Mumber of Insults per day: 0.02	Less than thou >	New Neighborhood Factor:	1.00	New Neighborhood Factor:	1.00		121 Help
Age Factor: 0.00 Gender Factor: 1.00 Bullying History Factor: 0.00 Mental Health History Factor: 0.00 Mental Health History Factor: 0.00 Past Disciplinary Issues Factor: 0.25 Race/Ethnicity Factor 0.00 Intermet Usage Factor: 0.25 Race/Ethnicity Factor 0.00 Potential Bullies: 2 Number of Days Scanned: 90 Number of Insults per day: 0.02 Image Accures Image Accures Image Accure		New School Factor:	1.00	New School Factor:	1.00	1	National Suisida Drevention
Gender Factor: 1.00 Bullying History Factor: 0.00 Mental Health History Factor: 0.00 Past Disciplinary Issues Factor: 0.25 Race/Ethnicity Factor: 0.00 Number of Days Scanned: 90 Number of Days Scanned: 90 Number of Insults per day: 0.02 Mumber of Insults per day: 0.02 Mumber of Insults per day: 0.02		Age Factor:	0.00	Age Factor:	1.00		National Suicide Prevention
Bullying History Factor: 0.00 Mental Health History Factor: 0.00 Mental Health History Factor: 0.00 Mental Health History Factor: 0.00 Internet Usage Factor: 0.00 Potential Bullies: 2 Number of Days Scanned: 90 Number of Insults per day: 0.02 Mumber of Ins		Gender Factor:	1.00	Gender Factor:	1.00	(APR)	A Thin Line
Mental Health History Factor: 0.00 Past Disciplinary Issues Factor: 0.00 Internet Usage Factor: 0.25 Race/Ethnicity Factor: 0.25 Race/Ethnicity Factor: 0.00 Potential Bullies: 2 Number of Days Scanned: 90 Number of Insults per day: 0.02 Mumber of Insults per day: 0.80 Mental Health History Factor: 1.00 Potential Bullies: 2 Number of Insults per day: 0.80 Mental Health Health History Factor: 1.00 BullyingRank Werther Mumber of Insults per day: 0.80 Mental Health Health History Factor: 1.00 Sexual Orientation Factor: 0.00 Potential Bullies: 2 Number of Insults per day: 0.80 Mental Health History Factor: 1.00 Sexual Orientation Factor: 0.00 Potential Bullies: 2 Number of Insults per day: 0.80 Mental Health History Factor: 1.00 Sexual Orientation Factor: 0.00 Potential Bullies: 2 Number of Insults per day: 0.80 Mental Health History Factor: 1.00 Sexual Orientation Factor: 0.00 Potential Bullies: 2 Number of Insults per day: 0.80 Mental Health History Factor: 1.00 Sexual Orientation Factor: 0.00 Potential Bullies: 2 Number of Insults per day: 0.80 Mental Health History Factor: 1.00 Sexual Orientation Factor: 0.00 Number of Insults per day: 0.80 Mental Health History Factor: 1.00 Number of Insults per day: 0.80 Mental Health Health History Factor: 1.00 Number of Insults per day: 0.80 Mental Health History Factor: 1.00 Mental Health History Factor: 0.25 Number		Bullying History Factor:	0.00	Bullying History Factor:	0.00		A Thin Line
Past Disciplinary Issues Factor: 0.00 Internet Usage Factor: 0.25 Race/Ethnicity Factor 0.00 Sexual Orientation Factor: 0.00 Potential Bullies: 2 Number of Days Scanned: 90 Number of Insults per day: 0.02 Mumber of Insults per day: 0.02 Mumber of Insults per day: 0.02 Mumber of Insults per day: 0.80 Mumber of Insults p		Mental Health History Factor:	0.00	Mental Health History Factor:	1.00	(AR)	Darp This Way Foundation
Internet Usage Factor: 0.25 Race/Ethnicity Factor 0.00 Sexual Orientation Factor: 0.00 Potential Bullies: 2 Number of Days Scanned: 90 Number of Insults per day: 0.02 $ \begin{array}{c} \hline \hline$		Past Disciplinary Issues Factor:	0.00	Past Disciplinary Issues Factor:	1.00		Born This way Foundation
Race/Ethnicity Factor 0.00 Sexual Orientation Factor: 0.00 Potential Bullies: 2 Number of Days Scanned: 90 Number of Insults per day: 0.02 Mumber of Insults per day: 0.80 Mumber of Insults per		Internet Usage Factor:	0.25	Internet Usage Factor:	0.25	(AR)	COMP
Sexual Orientation Factor: 0.00 Potential Bullies: 2 Number of Days Scanned: 90 Number of Insults per day: 0.02 Mumber of Insults per day: 0.80 Mumber of Insult		Race/Ethnicity Factor	0.00	Race/Ethnicity Factor	1.00		CSWD
Potential Bullies: 2 Number of Days Scanned: 90 Number of Insults per day: 0.02 Image: Potential Bullies: 2 Number of Days Scanned: 90 Number of Insults per day: 0.02 Potential Bullies: 2 Number of Days Scanned: 90 Number of Insults per day: 0.80 Image: Potential Bullies: 2 Number of Days Scanned: 90 Number of Insults per day: 0.80 Image: Potential Bullies: 2 Number of Days Scanned: 90 Number of Insults per day: 0.80 Image: Potential Bullies: 2 Number of Insults per day: 0.80 Image: Potential Bullies: 2 Number of Insults per day: 0.80 Image: Potential Bullies: 2 Number of Insults per day: 0.80 Image: Potential Bullies: 2 Number of Insults per day: 0.80 Image: Potential Bullies: 2 Number of Insults per day: 0.80 Image: Potential Bullies: 2 Number of Insults per day: 0.80 Image: Potential Bullies: 2 Number of Insults per day: 0.80 Image: Potential Bullies: 2 Number of Insults per day: 0.80 Image: Potential Bullies: 2 Number of Insults per day: 0.80 Image: Potential Bullies: 2 Number of Insults per day: 0.80 Image: Potential Bullies: 2 Number of Insults per day: 0.80 Image: Potential Bullies: 2 Number of Insults per day: 0.80 Image: Potential Bullies: 2 Number of Insults per day: 0.80 Image: Potential Bullies: 2 Number of Insults per day: 0.80 Image: Potential Bullies: 2 Number of Insults per day: 0.80 Image: Potential Bullies: 2 Number of Insults per day: 0.80 Image: Potential Bullies: 2 Number of Insults per day: 0.80 Image: Potential Bullies: 2 Number of Insults per day: 0.80 Image: Potential Bullies: 2 Number of Insults per day: 0.80 Image: Potential Bullies: 2 Number of Insults per day: 0.80 Ima		Sexual Orientation Factor:	0.00	Sexual Orientation Factor:	0.00	(AR)	Cubarbullying Org
Number of Days Scanned: 90 Number of Days Scanned: 90 Number of Insults per day: 0.02 Image: BullyingRank Image: BullyingRank Im		Potential Bullies:	2	Potential Bullies:	2		Cyberbullyling.Org
Number of Insults per day: 0.02 Image: BullyingRank Image: BullyingRank Image: BullyingRank 0.02 Image: BullyingRank Image		Number of Days Scanned:	90	Number of Days Scanned:	90		DARE
Image: BullyingRank Image: BullyingRank<		Number of Insults per day:	0.02	Number of Insults per day:	0.80		
Image: Sources Image						(##¥	Dion Initiative
BullyingRank User Profile Resources About Us BullyingRank User Profile Resources About Us BullyingRank User Profile Resources About Us	🗐 🜔 🖉 K		R		R		
	BullyingRank User Profile Resources About Us	BullyingRank User Profile Resources	About Us	BullyingRank User Profile Resources	About Us	BullyingRank	User Profile Resources At
					,	1	

Background

In response to the growing prevalence of cyberbullying – the deliberate use of online digital media to communicate false, embarrassing, or hostile information about another person – among teens, researchers from a range of disciplines – including psychology and computer science – have been working to develop a better understanding of how cyberbullying occurs, who is at increased risk, and what we can do to prevent it. The efforts of psychologists have revolved primarily around explanatory models that shed light on key predictors and outcomes of cyberbullying, whereas the efforts of computer scientists have been directed towards identification models that can detect cyberbullying using internet or social media content. There has been surprisingly little interdisciplinary research that bridges the findings and developments from these disparate fields.

Identification

The BullyBlocker app uses the Cyberbullying Identification Module to calculate the Bullying Rank for a teen. This module uses information from the Data Collection *Module* – including Facebook data and parent-user data – to identify any risk factors or warning signs that may be present in the teen's profile. Each risk factor and warning sign – from insulting wall posts and comments to the teen's gender and mental health history – has been demonstrated through social science research to predict cyberbullying risk. To that end, specific risk factors and warning signs have been assigned a "weight," or number that indicates how much influence it should have in the calculation of the Bullying Rank, using the effect sizes indicated in social science research as a guide.



Risk Factors

Risk factors currently incorporated into the calculation of the Bullying Rank include age, gender, past history of bullying, internalizing problems, externalizing problems, race, sexual orientation, and internet use. Most risk factors were empirically supported by meta-analyses in cyberbullying research – some are included as exploratory risk factors to determine if they are correlated with changes in cyberbullying risk.

To incorporate each risk factor, one of three chosen meta-analyses first identified the specific factor as having a significant relationship with a teen's cyberbullying experience. Risk factors were then evaluated for their applicability to the app – things such as ease of extractability from Facebook, or the likelihood of a parent's accurate report, were considered – before incorporating each risk factor into the identification module. Lastly, the meta-analytic effect size for each risk factor was used to assign an approximate weight for the risk factor within the final Bullying Rank calculation. See the associated table for a detailed description of the metaanalyses used and where each risk factor was found.

Factor	Meta-Analysis	Details	Weight
New School	*	# days in a new school	0.10
New Neighborhood	*	# days in a new neighborhood	0.10
Age	Kowalski et al. (2014)	Applied if value is 11-16	0.04
Gender	Guo (2016)	Applied if value is female	0.12
Race/ Ethnicity	**	Applied if race is non-White or if ethnicity is Hispanic/Latino	0.02
Sexual Orientation	Fedewa & Ahn (2011)	Applied if self-identified as LGBTQ	0.29
Past Bullying	Kowalski et al. (2014)	Applied if user experienced bullying in last 1 month, 1-2 months, more than 2 months	0.42
Daily Internet Use	Kowalski et al. (2014)	Considers ranges <1h, 1h-3h, 4h-6h, >6h	0.17
Internalizing Problems	Guo (2016)	Considers history of depression, anxiety, low self-esteem	0.28
Externalizing Problems	Guo (2016)	Considers history of disciplinary issues or substance use	0.21

*Risk factor is included in current model on an exploratory basis to investigate correlation with cyberbullying risk.

**Meta-analytic effect was found to be nonsignificant in a previous review, but is included in the current model on an exploratory basis.

Evaluation

To evaluate the BullyBlocker's ability to effectively detect instances of cyberbullying, a simulated social network was created based on actual data from Twitter (i.e., tweets including both ambiguous and cyberbullying messages), cyberbullying messages from well-known cyberbullying cases in the media, and randomly-generated social media data.

The resulting dataset was then: (1) processed through the BullyBlocker app, and (2) coded independently by two members of the research team (i.e., human coders), to estimate a Bullying Rank. The appgenerated and human estimates of Bullying Rank were then compared.

A high degree of convergence was found between the risk estimates generated by the BullyBlocker app and the human coders.

Discussion

The BullyBlocker app is an innovative development in interdisciplinary research. This mobile app is novel due to its:

- 1) Use of information from parents *and* teens to predict cyberbullying risk
- 2) Integration of empirical findings from psychology to improve identification of cyberbullying risk
- 3) Generation of a customized list of resources for parents based on unique aspects of the cyberbullying their teen may be experiencing

BullyBlocker is a critical step forward in identifying ways to combat cyberbullying with the potential to enact positive change at the level of individuals, families, communities, and society, at large.

Average Error for Different Weights of Warning Signs and Vulnerability Factors







