BullyBlocker

Designing and Implementing a Model to Identify Cyberbullying in Facebook

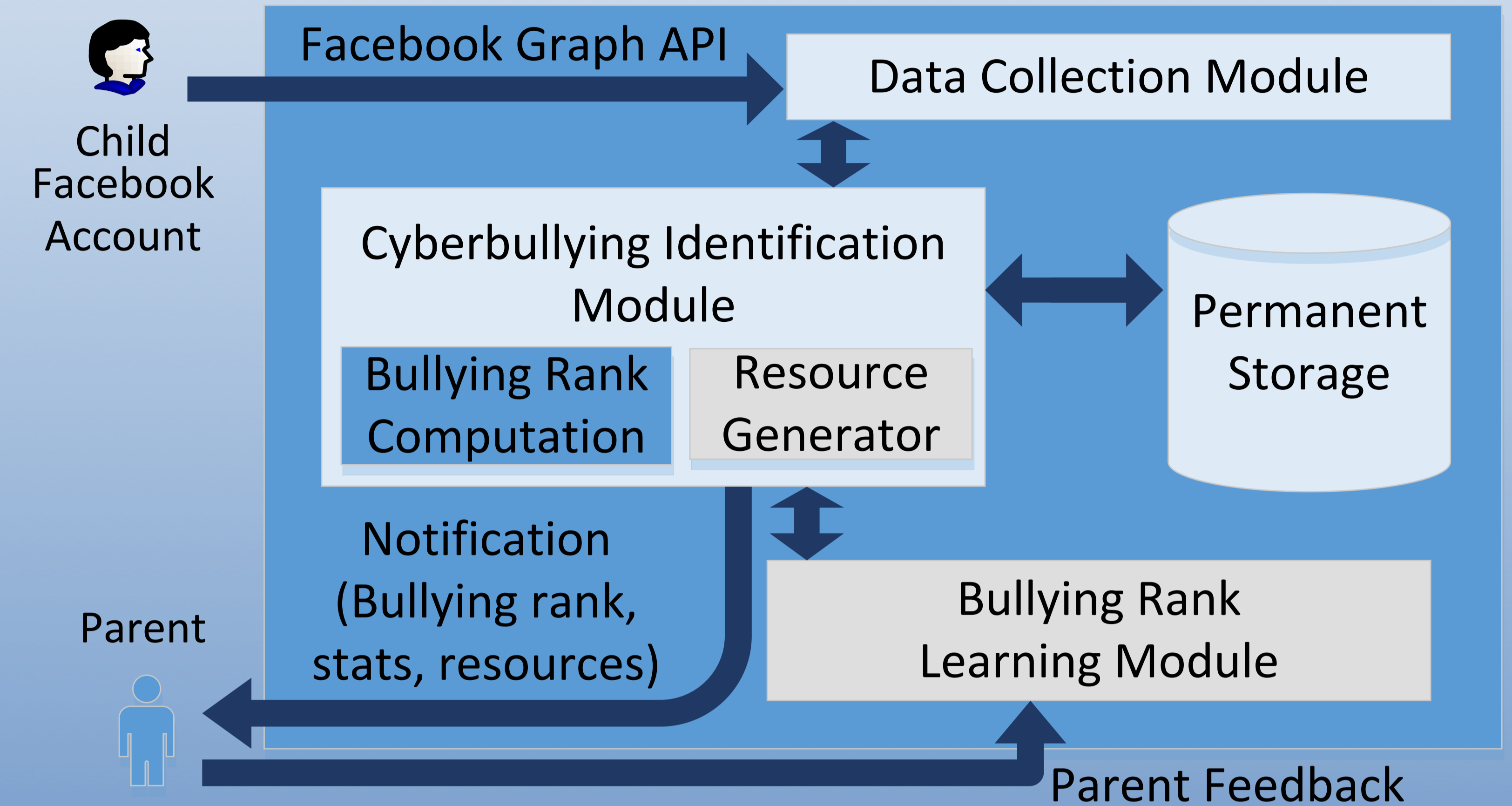Advisor: Yasin N. Silva   Students: Chance Brown | Liz Garcia | Bryan Sawkins

Problem and Contributions

- Cyberbullying is the use of online digital media to communicate false, embarrassing, or hostile information about another person.
- It is the most common online risk for adolescents and well over half of young people do not tell their parents when it occurs.
- There have been many studies about the nature and prevalence of cyberbullying.
- Relatively less work in the area of automated identification of cyberbullying.
- The focus of our work is to develop an automated model to identify and measure the degree of cyberbullying in social networking sites.
- We present the design of a model for identifying cyberbullying that builds on previous research findings in the areas of traditional bullying and cyberbullying in adolescents.
- We also identify challenges and opportunities to integrate the latest results from psychology and social network data analysis to address a problem of great social impact.
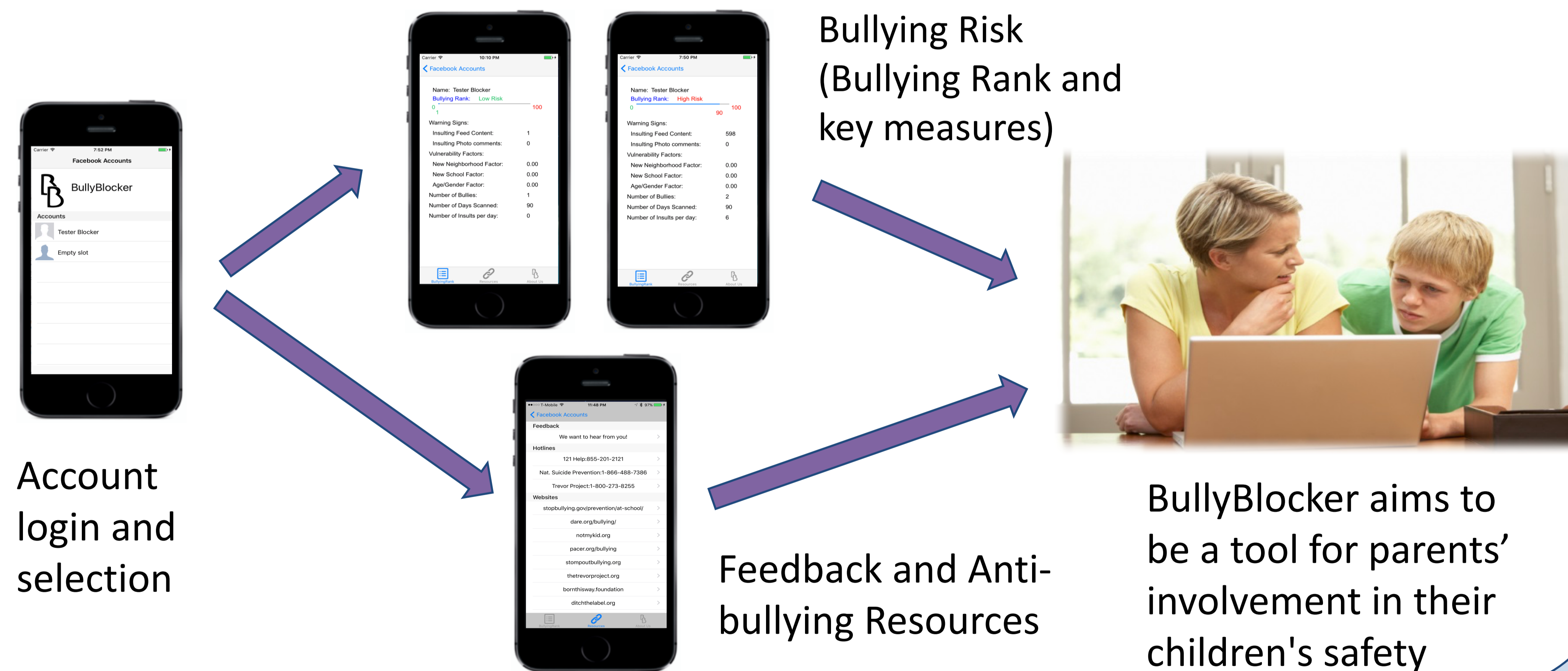
Design

- BullyBlocker analyzes adolescents' interactions with their social network to identify:
  - Warning signs: e.g., number of insulting messages and embarrassing pictures.
  - States of vulnerability: e.g., newcomers, age-gender group, members of minority groups, people with disabilities, etc.
- The integration of these factors is guided by previous research results in psychology.
- Bullying Rank: Estimates the probability of a minor experiencing cyberbullying.
- The computed Bullying Rank is returned to the parent or guardian of the minor.
- The Bullying Rank Is divided into three normalized levels of risk intensity: Low[0-33], Medium[34-66], High[67-100]
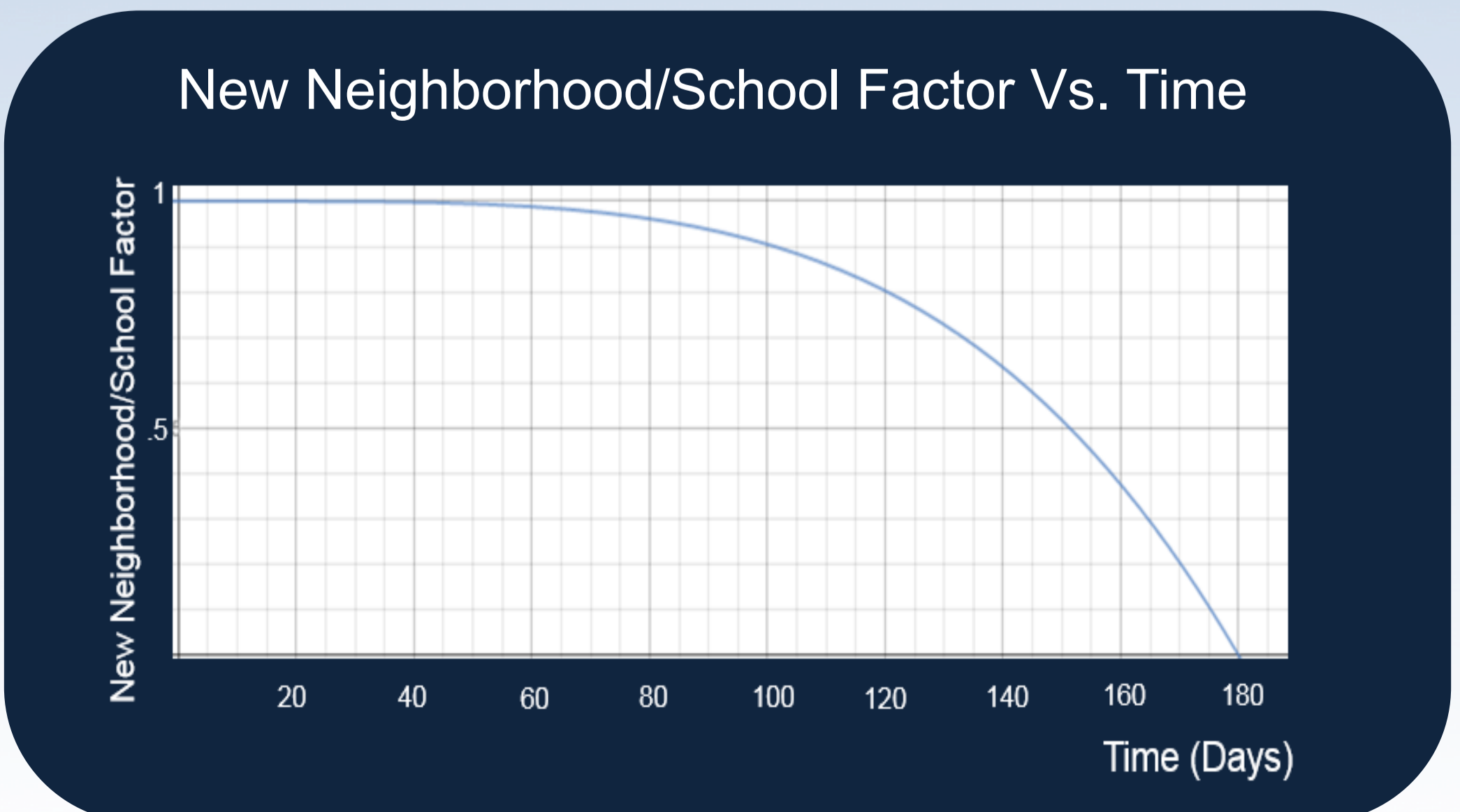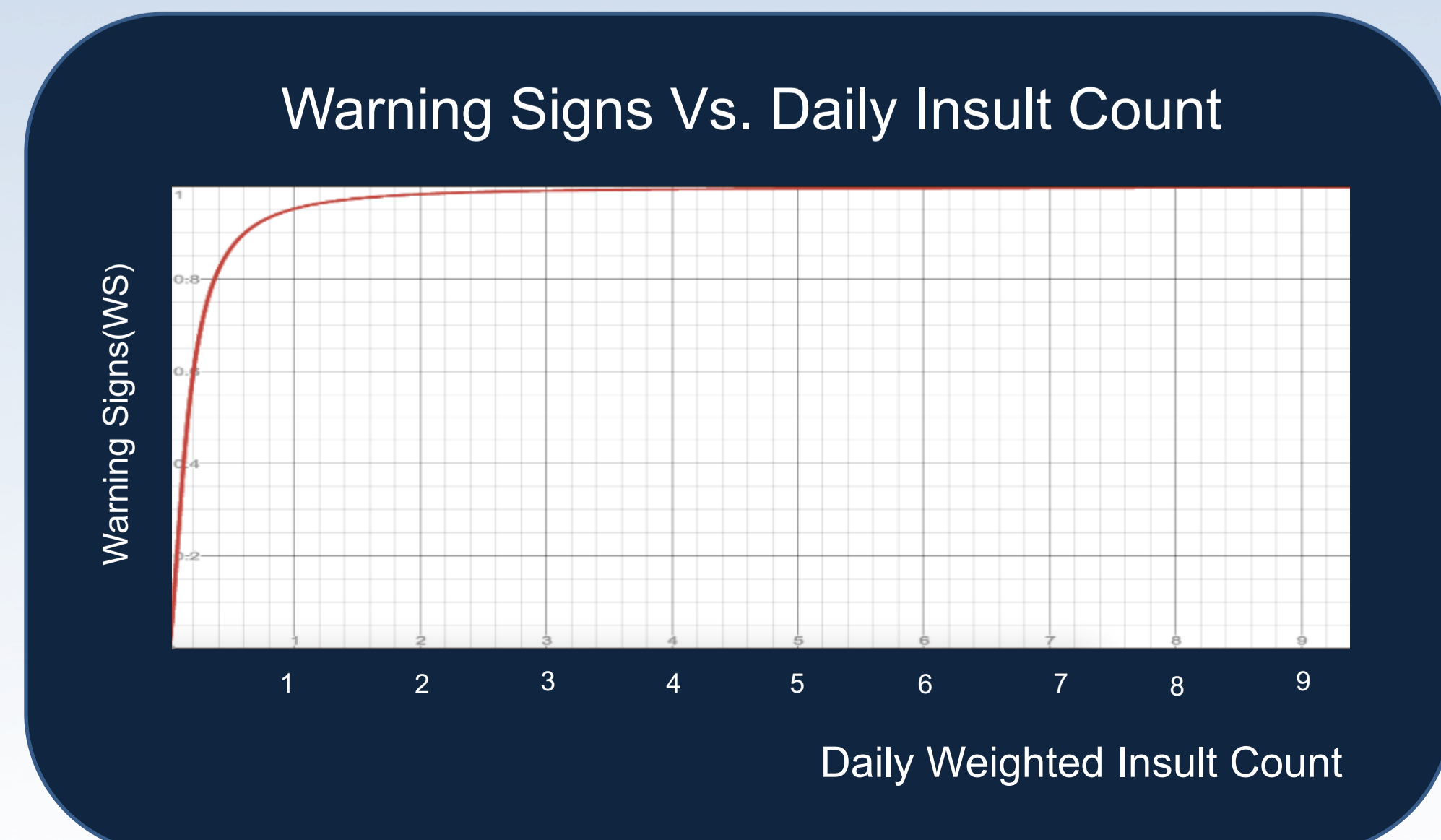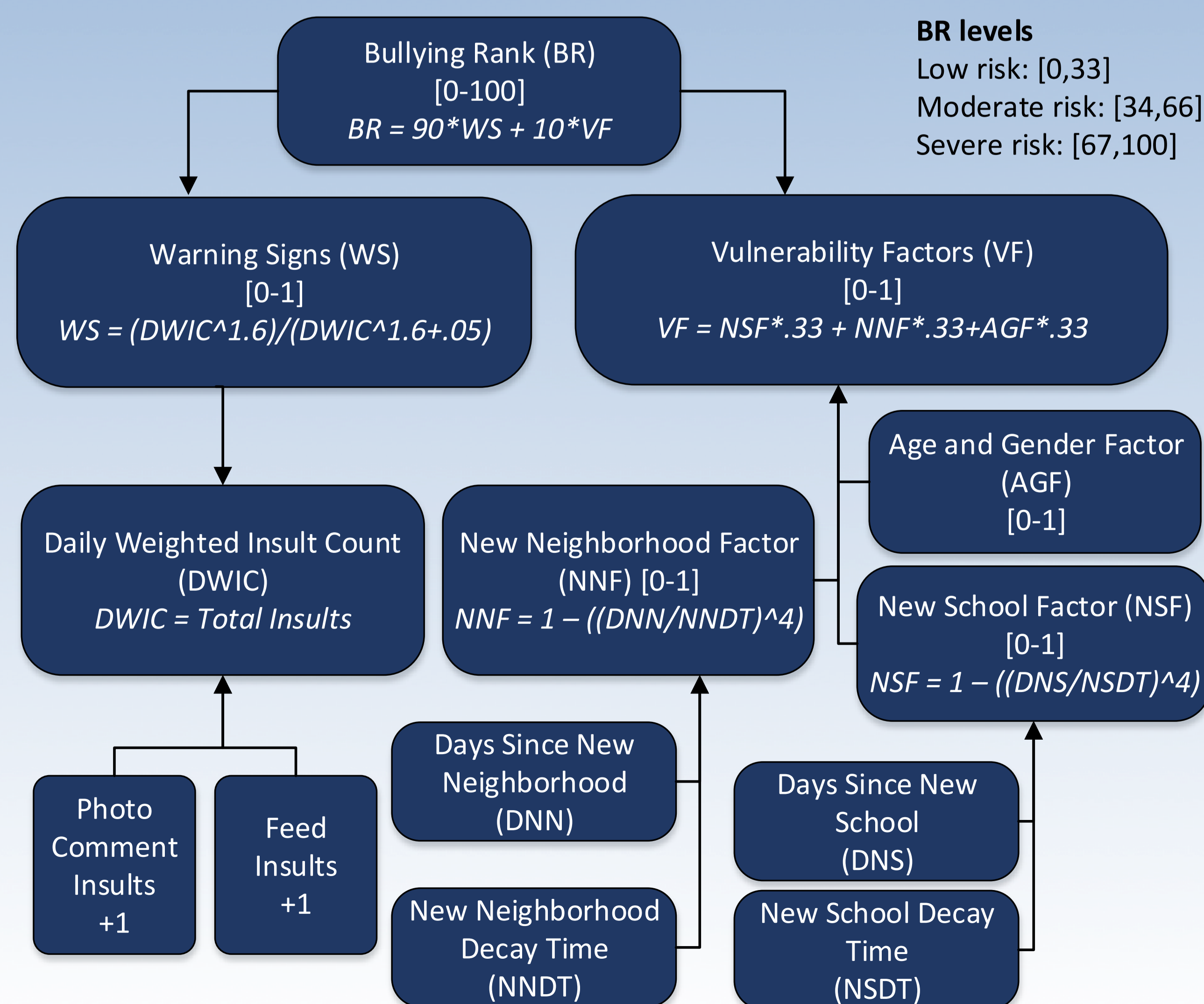
Architecture

Child Facebook Account — Facebook Graph API → Data Collection Module

Cyberbullying Identification Module
- Bullying Rank Computation
- Resource Generator

Permanent Storage

Notification (Bullying rank, stats, resources)

Bullying Rank Learning Module

Parent

Parent Feedback

Identification and Notification Process

Account login and selection

Bullying Risk (Bullying Rank and key measures)

Feedback and Anti-bullying Resources

BullyBlocker aims to be a tool for parents' involvement in their children's safety

Publications

- Y. N. Silva, C. Rich, J. Chon, L. M. Tsosie. *BullyBlocker: An App to Identify Cyberbullying in Facebook.* The 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), San Francisco, CA, USA, 2016.
- Y. N. Silva, C. Rich, D. Hall. *BullyBlocker: Towards the Identification of Cyberbullying in Social Networking Sites.* The 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), San Francisco, CA, USA, 2016.
- L. M. Tsosie, Y. N. Silva. *Facebully: Towards the Identification of Cyberbullying in Facebook.* The Grace Hopper Celebration of Women in Computing (GHC), Minnesota, USA, 2013.

Risk Factors

BR levels
Low risk: [0,33]
Moderate risk: [34,66]
Severe risk: [67,100]

Bullying Rank (BR) [0-100]
$$BR = 90*WS + 10*VF$$

Warning Signs (WS) [0-1]
$$WS = (DWIC^{1.6})/(DWIC^{1.6}+.05)$$

Vulnerability Factors (VF) [0-1]
$$VF = NSF*.33 + NNF*.33 + AGF*.33$$

Daily Weighted Insult Count (DWIC) [0-1]
$$DWIC = Total\ Insults$$

New Neighborhood Factor (NNF) [0-1]
$$NNF = 1 – ((DNN/NNDT)^4)$$

Age and Gender Factor (AGF) [0-1]

New School Factor (NSF) [0-1]
$$NSF = 1 – ((DNS/NSDT)^4)$$

Photo Comment Insults +1

Feed Insults +1

Days Since New Neighborhood (DNN)

New Neighborhood Decay Time (NNDT)

Days Since New School (DNS)

New School Decay Time (NSDT)

Warning Signs Vs. Daily Insult Count

Warning Signs(WS) vs Daily Weighted Insult Count

New Neighborhood/School Factor Vs. Time

New Neighborhood/School Factor vs Time (Days)