
PRELIMINARY VERSION

Harmony Institute Score (HIScore)

A Measure of A Movie's Buzz, Sentiment and Influence

`www.harmony-institute.org · info@harmony-institute.org`

Twitter Buzz

P_{t_d} : number of positive tweets about movie on day t_d

N_{t_d} : number of neutral tweets about movie on day t_d

T_{t_d} : total number of tweets (positive, neutral and negative) about movie on day t_d

I_{t_d} : number of followers of people who tweeted about movie on day t_d

T_{t_q} : total number of tweets (positive, neutral and negative) about movie in quarter t_q

I_{t_q} : number of followers of people who tweeted about movie in quarter t_q

$Twitter_{t_d}$: daily Twitter buzz on day t_d

$Twitter_{t_d}^s$: smoothed daily Twitter buzz on day t_d

x : number of days used in simple moving average computation for Twitter buzz

$$\begin{aligned}
 Twitter_{t_d} &= [(P_{t_d} + N_{t_d}) \frac{\log(I_{t_d})}{T_{t_d}} + 1] \\
 &\quad \times \underbrace{\sqrt{\frac{\frac{I_{t_q-4 \times 3}}{T_{t_q-4 \times 3}}}{\frac{I_{t_q}}{T_{t_q}}}}}_{\text{Adjustment for growth in Twitter userbase over time}} \\
 Twitter_{t_d}^s &= x\text{-day simple moving average of } Twitter_{t_d}
 \end{aligned}$$

where the time period, x , can be adjusted according to the degree of smoothness desired. Currently $x = 7$.

TV Buzz

M_{t_d} : number of times¹ movie mentioned in TV show(s) on day t_d

T_{t_d} : number of TV shows that mentioned movie on day t_d

TV_{t_d} : daily TV buzz on day t_d

$TV_{t_d}^s$: smoothed daily TV buzz on day t_d

x : number of days used in simple moving average computation for TV buzz

$$TV_{t_d} = \log(M_{t_d}) \times T_{t_d} + 1$$

$$TV_{t_d}^s = x\text{-day simple moving average of } TV_{t_d}$$

where the time period, x , can be adjusted according to the degree of smoothness desired. Currently $x = 14$.

¹If a movie is mentioned twice in a TV show and then once in another TV show on day t_d , $M_{t_d} = 3$

Newspaper Buzz

M_{t_d} : number of times² movie mentioned in newspaper article(s) on day t_d

T_{t_d} : number of newspaper articles that mentioned movie on day t_d

$News_{t_d}$: daily newspaper buzz on day t_d

$News_{t_d}^s$: smoothed daily newspaper buzz on day t_d

x : number of days used in simple moving average computation for newspaper buzz

$$News_{t_d} = \log(M_{t_d}) \times T_{t_d} + 1$$

$$News_{t_d}^s = x\text{-day simple moving average of } News_{t_d}$$

where the time period, x , can be adjusted according to the degree of smoothness desired. Currently $x = 14$.

²If a movie is mentioned twice in a newspaper article and then once in another newspaper article on day t_d , $M_{t_d} = 3$

Internet search buzz

G_{t_w} : weekly Google search activity³ on week t_w

G_{t_d} : daily Google search activity on day t_d

$Search_{t_d}^s$: smoothed daily Google search buzz on day t_d

x : number of days used in variance computation for Internet search buzz

$$G_{t_d} = G_{t_w} + 1 \quad \text{where } t_d \subset t_w$$
$$Search_{t_d}^s = \begin{cases} -\text{Variance}(G_{t_d-x}, \dots, G_{t_d}) & \text{if } G_{t_d} - G_{t_d-x} < 0 \\ \text{Variance}(G_{t_d-x}, \dots, G_{t_d}) & \text{otherwise} \end{cases}$$

where the time period, x , can be adjusted according to the degree of smoothness desired. Currently $x = 28$.

³Scaled by Google from 0 to 100 based on the range of search activity in the time period specified by user

Scaling

Daily Twitter buzz, TV buzz, newspaper buzz and Google search buzz have now been calculated. However, these four components may have different ranges. In order to combine these four components into one measure, they must first be scaled individually to a standard scale, chosen in this case to be 1-100.

For each of the four components, scaled buzz is computed by letting the minimum of each component be 1 and the maximum of each component be 100.

HIScore⁴

$Twitter_{t_d}^{ss}$: scaled, smoothed daily Twitter buzz on day t_d

$TV_{t_d}^{ss}$: scaled, smoothed daily TV buzz on day t_d

$News_{t_d}^{ss}$: scaled, smoothed daily newspaper buzz on day t_d

$Search_{t_d}^{ss}$: scaled, smoothed daily Google search buzz on day t_d

$HIScore_{t_d}^s$: daily HIScore on day t_d

$HIScore_{t_d}^{ss}$: smoothed daily HIScore on day t_d

$HIScore_{t_d}^{sss}$: locally weighed and smoothed daily HIScore on day t_d

w_a : weight of component a

s : smoothing factor used in locally weighted regression model

x : number of days used in simple moving average computation for newspaper buzz

$$\begin{aligned} HIScore_{t_d}^s &= w_{Twitter_{t_d}^{ss}} \times Twitter_{t_d}^{ss} + w_{TV_{t_d}^{ss}} \times TV_{t_d}^{ss} + w_{News_{t_d}^{ss}} \times News_{t_d}^{ss} + w_{Search_{t_d}^{ss}} \times Search_{t_d}^{ss} \\ HIScore_{t_d}^{ss} &= x\text{-day simple moving average of } HIScore_{t_d}^s \\ HIScore_{t_d}^{sss} &= \text{locally weighted regression model (lowess) of } HIScore_{t_d}^{ss} \text{ by time} \\ &\quad \text{with smoothing factor } s \end{aligned}$$

where the weights can be adjusted according to the relative importance of the four components, and the smoothing factor and time period can be adjusted according to the degree of smoothness desired. Currently, the parameters are as follows:

⁴On the scale of 1-100, the higher the better

$$w_{Twitter_{t_d}^{ss}} = 0.2$$

$$w_{TV_{t_d}^{ss}} = 0.3$$

$$w_{News_{t_d}^{ss}} = 0.3$$

$$w_{Search_{t_d}^{ss}} = 0.2$$

$$x = 15$$

$$s = 0.1$$