

Doing the math 'predicts' which movies will be box office hits

August 22 2013



Credit: Pixabay from Pexels

(Phys.org) —Researchers have devised a mathematical model which can be used to predict whether films will become blockbusters or flops at the box office – up to a month before the movie is released.

Their model is based on an analysis of the activity on Wikipedia pages about American films released in 2009 and 2010. They examined 312 [movies](#), taking into account the number of page views for the movie's article, the number of human editors contributing to the article, the number of edits made and the diversity of online users.

The researchers from Oxford University, the Central European University at Budapest, and Budapest University of Technology and Economics have published their findings in the journal *PLOS ONE*.

The model was applied retrospectively so the researchers systematically charted the online buzz on Wikipedia around particular films and compared this with the box office takings from the first weekend after release. The results of the comparison between the predicted opening weekend revenue, using their mathematical model, and the actual figures (published in Internet Movie Database [IMDb]) showed a high degree of [correlation](#).

Their [mathematical algorithm](#) allowed them to predict box office revenues with an overall accuracy of around 77%. The study authors say this level of accuracy is higher than the best existing [predictive models](#) applied by marketing firms (which they estimate to be at around 57%). They could predict the box office takings of six out of 312 films with 99% accuracy where the predicted value was within 1% of the real value. Some 23 movies were predicted with 90% accuracy and 70 movies with an accuracy of 70% and above.

The more successful the film, the more accurately the researchers were able to predict box office takings. In the study, they explain that this is possibly due to the increased amount of online data generated by films that turn out to be successes. The model correctly forecast the commercial success of Iron Man 2, Alice in Wonderland, Toy Story 3 and Inception, but failed to accurately forecast the financial return on

the less successful movies *Never Let Me Go* and *Animal Kingdom*.

Dr Taha Yasseri, from the Oxford Internet Institute at the University of Oxford, said: 'These results can be of great value to marketing firms but more importantly for us, we were able to demonstrate how we can use socially generated online data to predict a lot about future human behaviour. The predicting power of the Wikipedia-based model, despite its simplicity compared with Twitter, is that many of the editors of the Wikipedia pages about the movies are committed movie-goers who gather and edit relevant material well before the release date. By contrast, the "mass" production of tweets occurs very close to the release time, and often these can be spun by marketing agencies rather than reflecting the feelings of the public.'

Co-author Professor János Kertész, from the Central European University of Budapest, Hungary, said: 'We have demonstrated for the first time that Wikipedia edit statistics provide us with another tool to predict social events. We studied the problem of predicting the financial success of movies and concluded that, in some aspects, forecasting based on Wikipedia outperforms tweets as Wikipedia activity has a longer timescale which enables earlier predictions.'

The study suggests that the efficiency of the predictions might be improved by applying more sophisticated statistical methods, such as including the controversy measure of an article. The [mathematical model](#) has not been applied yet to films that are not on release.

Provided by Oxford University

Citation: Doing the math 'predicts' which movies will be box office hits (2013, August 22) retrieved 19 April 2024 from <https://phys.org/news/2013-08-math-movies-office.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.