Visual Detection of Singularities in Recommendation Systems

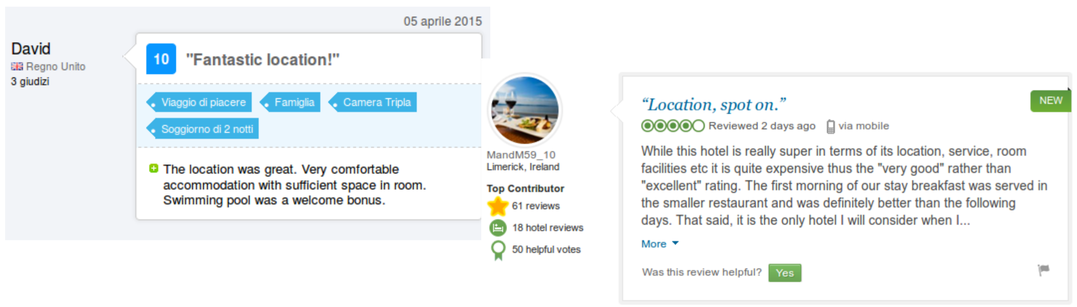
SAC 2015
Multimedia & Visualization
Salamanca, Spain
April 13-17, 2015

A. Colantonio[1], R. Di Pietro[2], M. Petrocchi[3], A. Spognardi[3]

[1]*Bay31 AG, Switzerland* [2]*Bell Labs Alcatel Lucent, Paris, France* [3]*CNR IIT, Pisa, Italy*

Introduction

Reviews can largely influence purchase decisions: if most reviews are positive, people are inclined to buy the product; on the contrary, negative reviews will persuade the user to look for alternatives. This fact leads to the **submission of fake reviews** with the main intent of twisting product perception [2]. Academia has shown a growing interest towards recommendation systems and a flourishing literature exists on deceptive reviews [4].



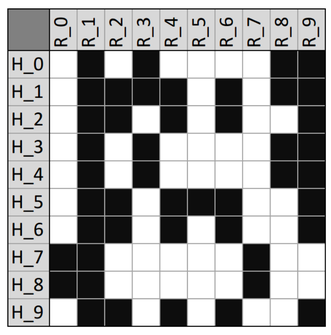
We propose analyzing reviews resorting to *visual analytics*—the science of analytic reasoning facilitated by visual interfaces [3]. This **strengthens human interaction with complex graphical displays**, putting humans in the loop allows for a **better correlation of knowledge with review data**. Moreover, visualization enables the analyst to visually mine data by means of human reasoning.

We propose a **visual approach** for singularity detection in review platforms. Our idea is to complement existing review assessment techniques by providing a "visual workbench" for review analysts. We adapted results coming from IT security [1] to implement a novel methodology for **review analysis**.

Methods

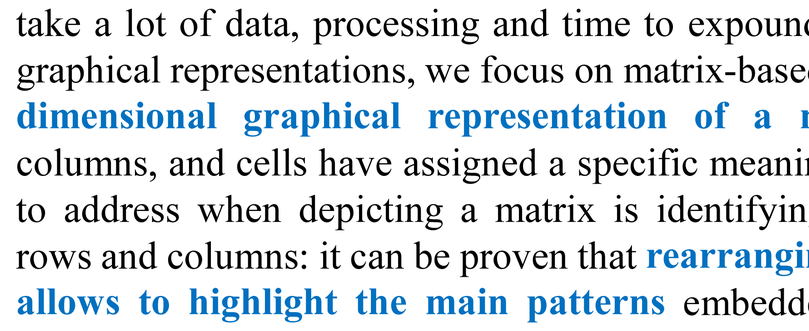
The problem that we address is offering a graphical way to effectively navigate existing reviews in a system, showing at a glance what it would take a lot of data, processing and time to expound. Among all possible graphical representations, we focus on matrix-based visualization, a **two-dimensional graphical representation of a matrix**, where rows, columns, and cells have assigned a specific meaning. The main problem to address when depicting a matrix is identifying a proper sorting of rows and columns: it can be proven that **rearranging rows and columns allows to highlight the main patterns** embedded in the represented data. To identify the best row/column sorting, we use a fast algorithm referred to as ADVISER (Access Data VISualizER) [1], which is able to provide a compact representation of patterns embedded in a binary matrix.

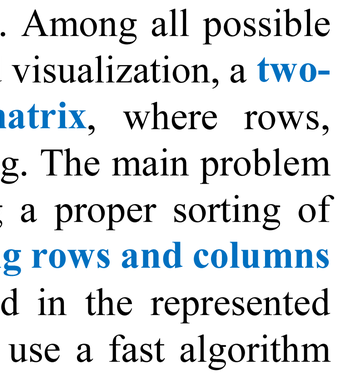
Hotel ratings
$\{\langle h_0, r_1\rangle, \langle h_0, r_3\rangle, \langle h_0, r_8\rangle, \langle h_0, r_9\rangle, \langle h_1, r_1\rangle, \langle h_1, r_2\rangle, \langle h_1, r_3\rangle, \langle h_1, r_4\rangle, \langle h_1, r_6\rangle, \langle h_1, r_8\rangle, \langle h_1, r_9\rangle, \langle h_2, r_1\rangle, \langle h_2, r_2\rangle, \langle h_2, r_4\rangle, \langle h_2, r_6\rangle, \langle h_2, r_9\rangle, \langle h_3, r_1\rangle, \langle h_3, r_3\rangle, \langle h_3, r_8\rangle, \langle h_3, r_9\rangle, \langle h_4, r_1\rangle, \langle h_4, r_3\rangle, \langle h_4, r_8\rangle, \langle h_4, r_9\rangle, \langle h_5, r_1\rangle, \langle h_5, r_2\rangle, \langle h_5, r_4\rangle, \langle h_5, r_5\rangle, \langle h_5, r_6\rangle, \langle h_5, r_9\rangle, \langle h_6, r_1\rangle, \langle h_6, r_2\rangle, \langle h_6, r_4\rangle, \langle h_6, r_6\rangle, \langle h_6, r_9\rangle, \langle h_7, r_0\rangle, \langle h_7, r_1\rangle, \langle h_7, r_7\rangle, \langle h_8, r_0\rangle, \langle h_8, r_1\rangle, \langle h_8, r_7\rangle, \langle h_9, r_1\rangle, \langle h_9, r_2\rangle, \langle h_9, r_4\rangle, \langle h_9, r_6\rangle, \langle h_9, r_9\rangle\}$



ADVISER algorithm

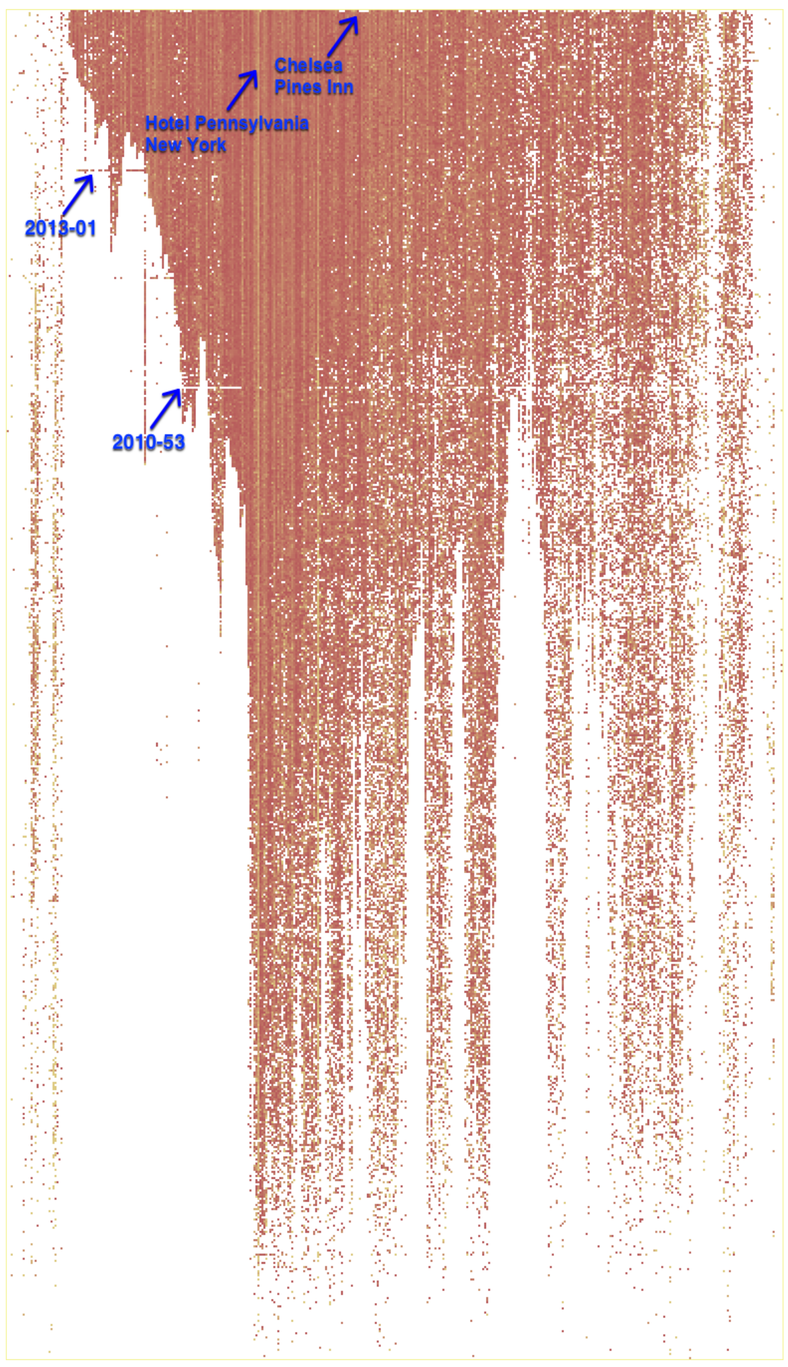
Cluster	Ratings	Hotels
c_1	$\{r_1\}$	$\{h_0, h_1, h_2, h_3, h_4, h_5, h_6, h_7, h_8, h_9\}$
c_2	$\{r_2, r_4, r_6, r_9\}$	$\{h_1, h_2, h_5, h_6, h_9\}$
c_3	$\{r_3, r_8, r_9\}$	$\{h_0, h_1, h_3, h_4\}$
c_4	$\{r_0, r_7\}$	$\{h_7, h_8\}$
c_5	$\{r_5\}$	$\{h_5\}$



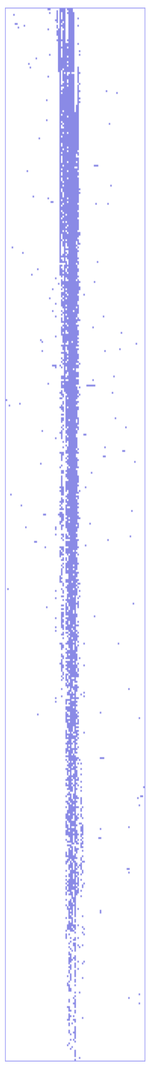


Our approach starts from specifying how to construct a matrix-based visualization out of review data. To this aim, we extracted real data from Tripadvisor (http://www.tripadvisor.com) and Booking.com (http://www.booking.com) websites. The extracted dataset consists of a list of reviews, each one made up of: user ID, hotel name, review date, review text, rating. We discuss the matrix representation obtained by assigning **each row to review weeks**, sorted from top to bottom in reverse chronological order; **columns represent hotels**, sorted according to the above mentioned ADVISER algorithm; finally, a cell is "filled" when there is at least one review for the given hotel in the given week.

Results

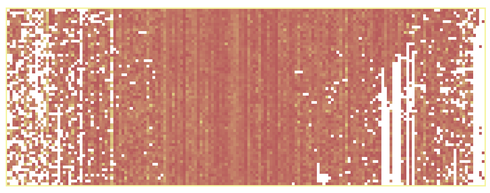


Tripadvisor reviews from August 27, 2001 to June 25, 2014 for hotels in New York. Each row of the matrix represents a week, sorted from top to bottom in reverse chronological order.
Yellow cell: given week average rating **1 star** (the lowest)
Red cell: given week average rating **5 stars** (the highest)



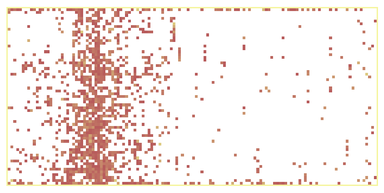
Tripadvisor reviews removed after September 10, 2013

Shared reviews between Tripadvisor and Booking.com

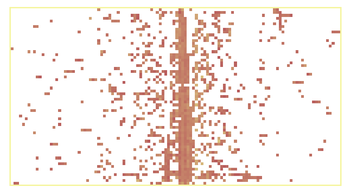


Reviews in Tripadvisor for hotels in common between the two platforms

Reviews in Booking.com for hotels in common between the two platforms



Reviews in Tripadvisor and *not* in Booking.com



Reviews in Booking.com and *not* in Tripadvisor

Conclusions

We have highlighted how to interpret raw review data as input to binary matrices that are the basis for the visualization and we have showed the effectiveness and quality of the proposed solution via an extensive experimental campaign over real data extracted from Tripadvisor and Booking.com. While the proposed approach is not a definitive answer to the vexed issue of detecting misleading reviews, we believe that our novel approach paves the way for further contributions in the area of visual analytics applied to recommendation systems.

Bibliography

[1] A. Colantonio, R. Di Pietro, A. Ocello, and N. V. Verde. Visual role mining: A picture is worth a thousand roles. IEEE Trans on Know. Data Eng., 24(6), 2012.
[2] B. Liu. Sentiment Analysis and Opinion Mining. Synthesis Lectures on Human Language Technologies. Morgan & Claypool Publishers, 2012.
[3] J. J. Thomas and K. A. Cook. A visual analytics agenda. Computer Graphics and Applications, IEEE, 26(1), 2006.
[4] Z. Xie and S. Zhu. Grouptie: toward hidden collusion group discovery in app stores. In WiSec'14, 2014.

Acknowledgments

This research has been partially supported by the MIB (My Information Bubble) project of *Registro.it.*