Behavioral pattern mining in web based educational system

Sanda Dragoş, **Diana Haliţă** and Christian Săcărea

Babeş-Bolyai University, Department of Computer Science, Cluj-Napoca, România
sanda@cs.ubbcluj.ro, **diana.halita@ubbcluj.ro**, csacarea@math.ubbcluj.ro

September 2015

Prerequisites | Quantitative View on Performance Evaluation | Qualitative View on Performance Evaluation | Conclusions and Future

oooo | oooooooooooooo | oooooooooooooo | oo

TABLE OF CONTENTS

WHAT WE NEEDED...

- understand visitors' behavior while using e-learning platforms;
- improving e-learning instrument's design according to users' needs;
- discover behavioral patterns.

OUR APPROACH...

■ a comprehensive step-wise approach:
  □ performance evaluation by means of some web analytics tools;
  □ visualizing data structure with force directed graphs;
  □ knowledge discovery using conceptual landscapes;
  □ triadic approach over behavioral categories.

Prerequisites   Quantitative View on Performance Evaluation   Qualitative View on Performance Evaluation   Conclusions and Future

○○●○   ○○○○○○○○○○○○   ○○○○○○○○○○○○   ○○

DATA PREPROCESSING

■ multiple sources, different formats, different granularities,;

■ = clean, transform, integrate data from other sources;

■ identification of sessions and users, the length of practice session and time to learn.

E-LEARNING INSTRUMENT

- **e-learning portal**: PULSE;
- **data collection**: February - July 2013;
- **investigate**: patterns of web usage behavior within PULSE.

Prerequisites   Quantitative View on Performance Evaluation   Qualitative View on Performance Evaluation   Conclusions and Future

○○○○   ●○○○○○○○○○○○○   ○○○○○○○○○○○○○   ○○

QUANTITATIVE WEB ANALYTICS ON PULSE

■ WATEC = on-site PHP web statistics tool:
  ☐ traffic graphics;
  ☐ tops ranked by the number of page views;
  ☐ details about location from where the clients access PULSE;
  ☐ details about operating systems and screen resolution;
  ☐ details about referrers.

■ offers an initial, minimal view on the entire set of data.

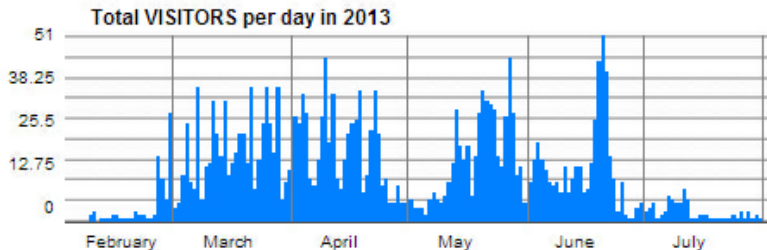# QUANTITATIVE WEB ANALYTICS ON PULSE

Figure 1: Daily visitors on PULSE

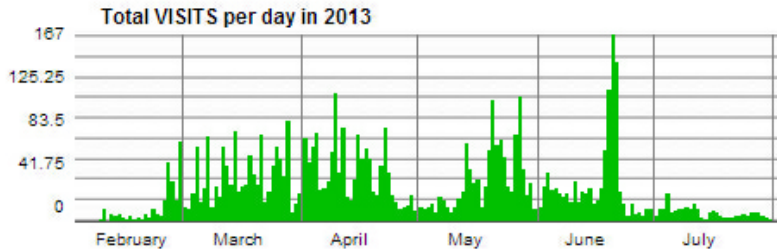# QUANTITATIVE WEB ANALYTICS ON PULSE

Figure 2: Daily visits on PULSE

QUANTITATIVE WEB ANALYTICS ON PULSE

- **confirmation**: students use PULSE and their activity is more intense during laboratory assignment deadlines and examinations;
- **view**: a quantitative overview on how many students access the site per day in various periods of year;
- **our benefit**: adjust the physical requirements of bandwidth and distributing the web server load onto multiple servers for the peak periods.

INTEGRATED VERSUS THIRD PARTY ANALYSIS

- **WATEC**:
  - □ sessions = actual HTTP sessions;
  - □ users = login ID of the user $\Rightarrow$ better identify individuals;;
  - □ eliminates the non authenticated visits.
- **AIU**:
  - □ users = distinct pairs of IP and User Agent;
- **GAL**:
  - □ users = cookies;

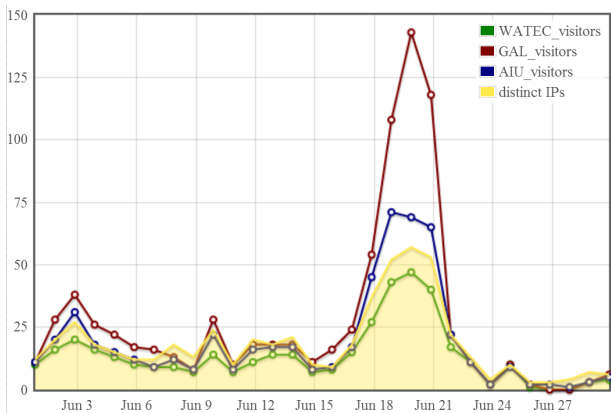# INTEGRATED VERSUS THIRD PARTY ANALYSIS

Figure 3: WATEC, GAL, AIU visitors - comparative results
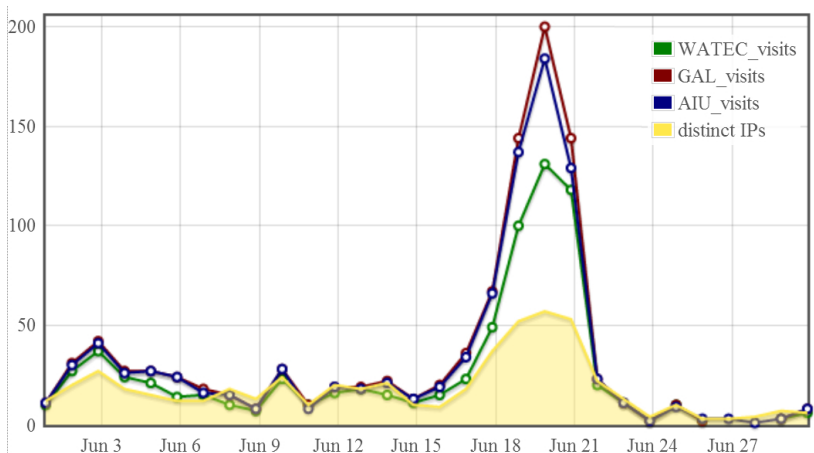
INTEGRATED VERSUS THIRD PARTY ANALYSIS

Figure 4: WATEC, GAL, AIU visits - comparative results

INTEGRATED VERSUS THIRD PARTY ANALYSIS

WATEC vs. AIU vs. GAL

Differences due to the fact that a student may access the website from multiple locations.

Prerequisites   Quantitative View on Performance Evaluation   Qualitative View on Performance Evaluation   Conclusions and Future

oooo   ooooooooo●oooo   oooooooooooooo   oo

INTEGRATED VERSUS THIRD PARTY ANALYSIS

- **confirmation**: the importance of using an "in-situ" evaluation instrument for the e-learning instrument;

- **view**: a comparative quantitative view between the three instruments in terms of the number of visitors / visits;

- **our benefit**: it offers most accurate information and therefore provides the most precise feedback.

Prerequisites | Quantitative View on Performance Evaluation | Qualitative View on Performance Evaluation | Conclusions and Future

○○○○ | ○○○○○○○○○●○○○ | ○○○○○○○○○○○○○ | ○○

FORCE DIRECTED GRAPHS ON PULSE DATA

- **purpose**: visualize data structured as objects and the relationship between them;
- **nodes**: classes of access files;
- **edges**: the trajectory used by users while navigation from a cluster of pages (one node) to another cluster of pages (another node);
- **weight**: the number of accesses that were made between the two nodes.
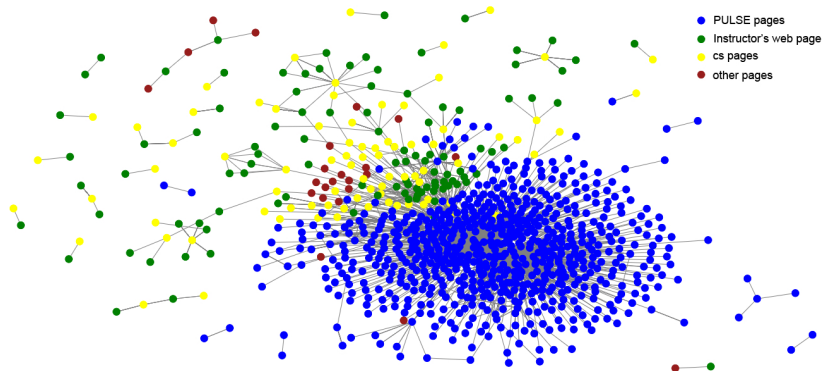
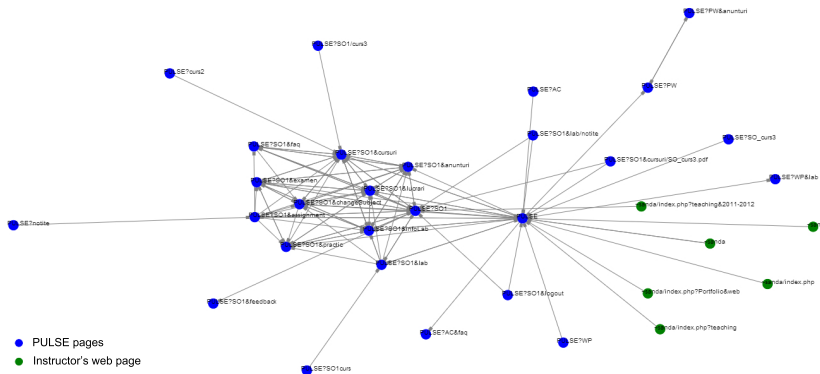# FORCE DIRECTED GRAPHS ON PULSE DATA

Figure 5: PULSE usage

Prerequisites   **Quantitative View on Performance Evaluation**   Qualitative View on Performance Evaluation   Conclusions and Future

○○○○   ○○○○○○○○○○○●○   ○○○○○○○○○○○○   ○○

FORCE DIRECTED GRAPHS ON PULSE DATA

Figure 6: PULSE used by students

Prerequisites   Quantitative View on Performance Evaluation   Qualitative View on Performance Evaluation   Conclusions and Future

oooo   ooooooooooooo●   oooooooooooooo   oo

FORCE DIRECTED GRAPHS ON PULSE DATA

- **confirmation**: students visit more often a nucleus of pages within PULSE which are closely connected;
- **view**: a quantitative view over PULSE accesses; if two nodes are closer to each other it means that users navigated more often between them;
- **our benefit**: rethink the structure of PULSE.

Prerequisites   Quantitative View on Performance Evaluation   **Qualitative View on Performance Evaluation**   Conclusions and Future

○○○○   ○○○○○○○○○○○○○   ●○○○○○○○○○○○○   ○○

QUALITATIVE ANALYSIS USING FORMAL CONCEPT ANALYSIS

- the data table from real world applications where objects are related to their properties by an incidence relation;

- knowledge is extracted from the data set for knowledge discovery, rule mining, other mining methodologies, exploration, knowledge aquisition and navigation among different concepts;

- Elba & ToscanaJ enabled a systematic exploration of the data set in order to detect behavioral patterns;

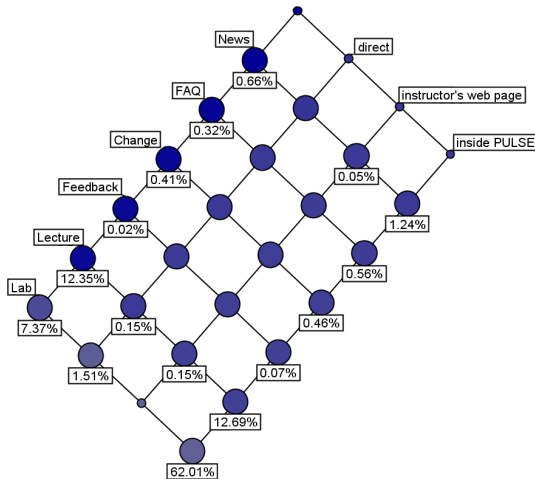# QUALITATIVE ANALYSIS USING FORMAL CONCEPT ANALYSIS

Figure 7: Access file classes visited before practical exam

Prerequisites   Quantitative View on Performance Evaluation   **Qualitative View on Performance Evaluation**   Conclusions and Future

oooo   oooooooooooooo   oo●oooooooooo   oo

QUALITATIVE ANALYSIS USING FORMAL CONCEPT ANALYSIS

- the referrer - referred pairs have been represented and visualized by ToscanaJ using a grid scale;

- ⇒ big interest for the LAB page (over 70%), given the specific of the time period that the scale has been built;

- some students access PULSE directly;

- **remark**: it is not advisable to change URI's too often.

QUALITATIVE ANALYSIS USING FORMAL CONCEPT ANALYSIS
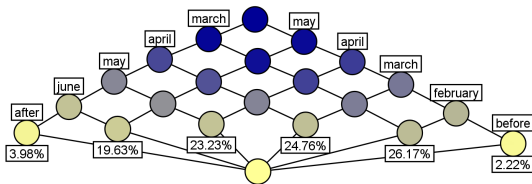
Figure 8: Interordinal scale of time intervals in the usage of PULSE

Prerequisites   Quantitative View on Performance Evaluation   **Qualitative View on Performance Evaluation**   Conclusions and Future

oooo   oooooooooooo   oooo●ooooooo   oo

QUALITATIVE ANALYSIS USING FORMAL CONCEPT ANALYSIS

■ the percentage of accesses for each month of the semester which are approximately uniform distributed;

■ **advantage of using scales**: scale choice, node zoom-in or zoom out, nested views, combining more scales into one browsing scenario, navigation between different concepts;

Prerequisites   Quantitative View on Performance Evaluation   **Qualitative View on Performance Evaluation**   Conclusions and Future

oooo   ooooooooooooo   ooooo●ooooooo   oo

QUALITATIVE ANALYSIS USING FORMAL CONCEPT ANALYSIS
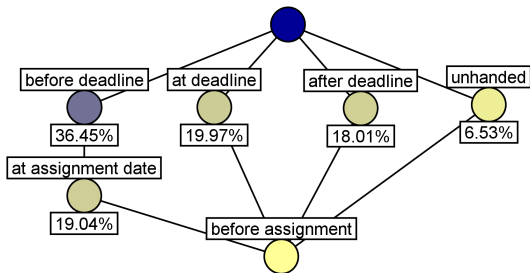
Figure 9: Handing L3 assignments
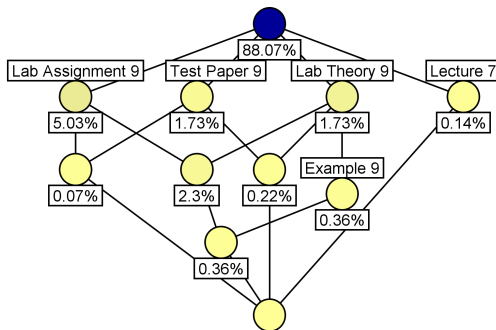
QUALITATIVE ANALYSIS USING FORMAL CONCEPT ANALYSIS

Figure 10: Related teaching support pages which students should visit for the 9th laboratory
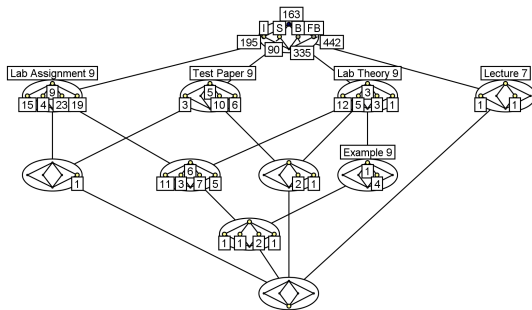
QUALITATIVE ANALYSIS USING FORMAL CONCEPT ANALYSIS

Figure 11: Related teaching support pages which students should visit for the 9th laboratory: Laboratory support (assignment, theory, example)

QUALITATIVE ANALYSIS USING FORMAL CONCEPT ANALYSIS

Figure 12: Related teaching support pages which students should visit for the 9th laboratory correlated with final marks

Prerequisites   Quantitative View on Performance Evaluation   **Qualitative View on Performance Evaluation**   Conclusions and Future

○○○○   ○○○○○○○○○○○○   ○○○○○○○○○●○○○   ○○

WEB USAGE MINING WITH TRIADIC FORMAL CONCEPT ANALYSIS

- **motivation**: different situations demand a third dimension for the underlying data structure;
- objects are related to attributes under some condition by a ternary incidence relation;
- **our triadic data set**: the pairs Referrer class-Access File class as attribute set, timestamps as conditions and students Login as the object set.

WEB USAGE MINING WITH TRIADIC FORMAL CONCEPT ANALYSIS
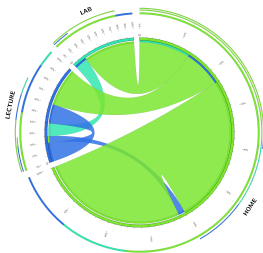
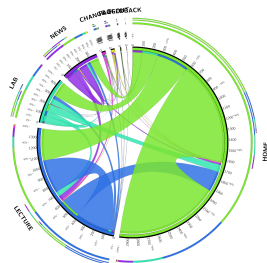Figure 13: SO: relaxed behavior
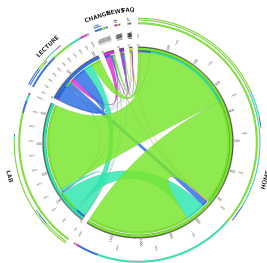
Figure 14: SO: normal behavior

Figure 15: SO: intense behavior

Prerequisites   Quantitative View on Performance Evaluation   **Qualitative View on Performance Evaluation**   Conclusions and Future

○○○○   ○○○○○○○○○○○○○   ○○○○○○○○○○○○●○   ○○

WEB USAGE MINING WITH TRIADIC FORMAL CONCEPT ANALYSIS

- the **relaxed** behaviour:
  - occurs mainly during holiday;
  - **remark**: fewer accesses;
  - HOME ⇒ LAB ⇒ LECTURE ⇒ HOME;

- the **normal** behaviour:
  - occurs during the semester;
  - **remark**: all classes are visited, especially LAB class;

- the **intense** behaviour:
  - occurs during examination periods;
  - **remark**: increased number of accesses, especially LECTURE class; HOME represents a connection to the other PULSE facilities;

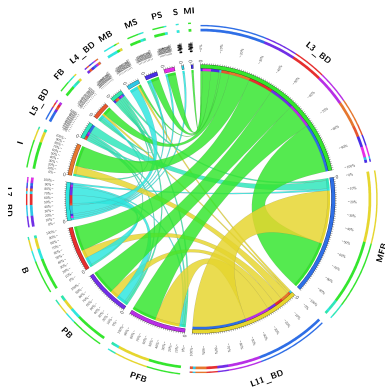# WEB USAGE MINING WITH TRIADIC FORMAL CONCEPT ANALYSIS

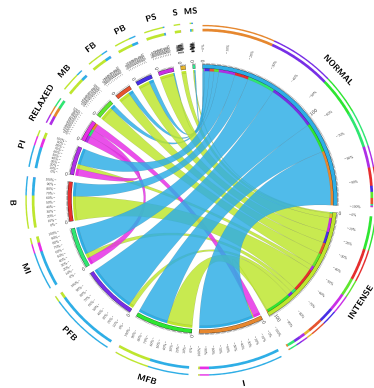Figure 16: Correlation between grades and before deadline

Figure 17: Correlation diagram between grades and behavior

CONCLUSIONS AND FUTURE WORK

- the **quantitative view** offers results from statistical analysis;

- scaled or circular visualizations provide a more **qualitative view** on the navigational pattern, comprising more details about how and where students navigate;

- **ADVANTAGE**:
  - the completeness of the information clustered in a concept.

- **FURTHER RESEARCH**:
  - defining so-called **attractors**, i.e. types of ideal behaviors to which users adhere over time by using the e-learning system;
  - detecting trend-setters, i.e. students which have successfully used the platform and hence encourage other students to follow them

Questions?