Association Rules
Visualization Related papers Summary
Sep 28, 2016

Summary

- 2D Matrices showing item to item relationship
- 3D Matrices Showing item to rule relationship
- Directed graphs, where direction shows Antecedent to Consequent Relationship
- Parallel Coordinates (PC)
- Combination of Graph and PC
- Trees

2-D Matrices : item to item association

Figure 2: Very difficult to determine the differences between (A+B→C) and (A→C and B→C).

Figure 1: The colored column indicates the association rule B→C. Different icon colors are used to show different metadata values of the association rule.

Graph Based

The direction of arrow shows the association from Antecedent to consequent

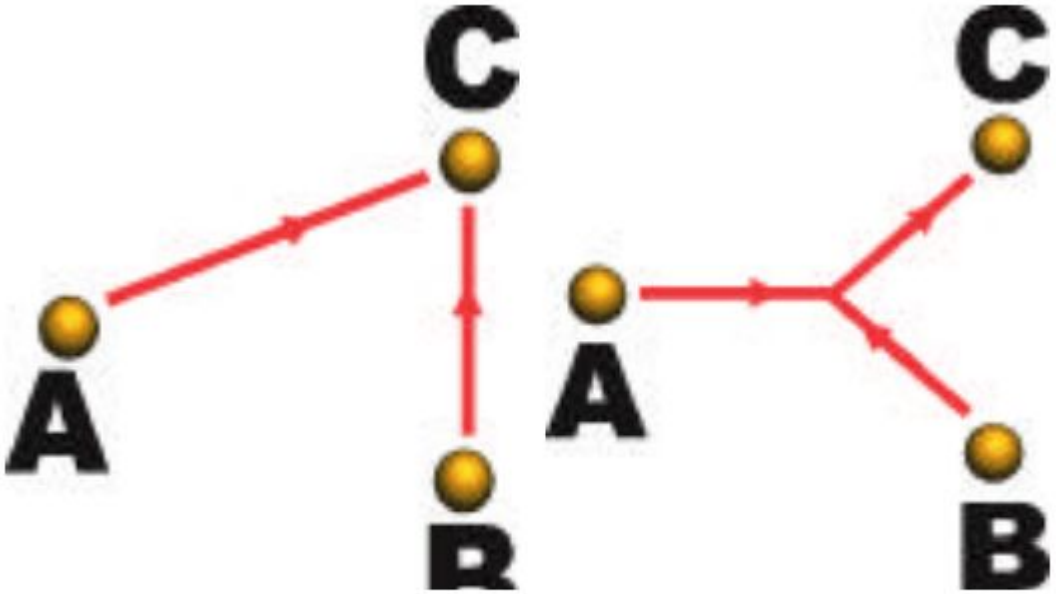


Figure 5: Left: A→C and B→C. Right: A+B →C.

3-D Matrix: rule to item

- Row shows a rule
- Where red color is antecedent and blue is consequent

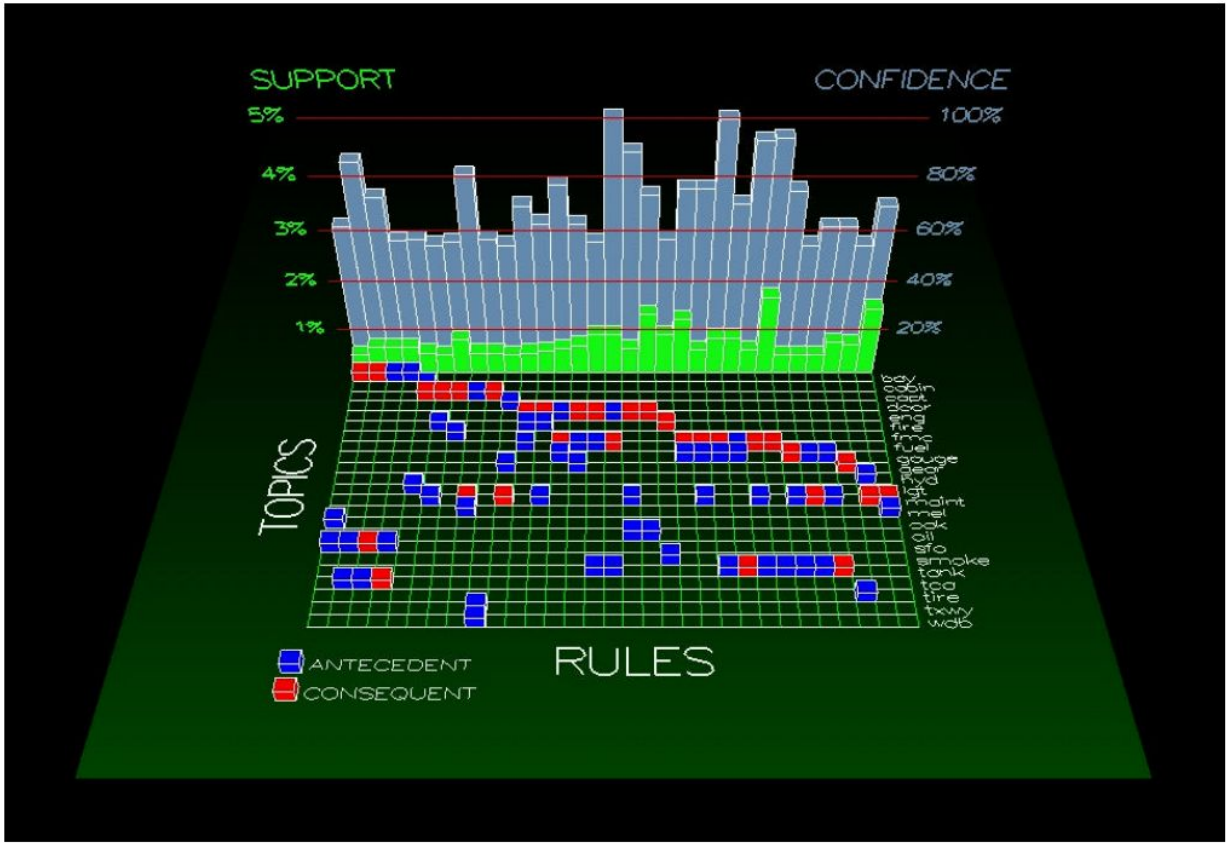
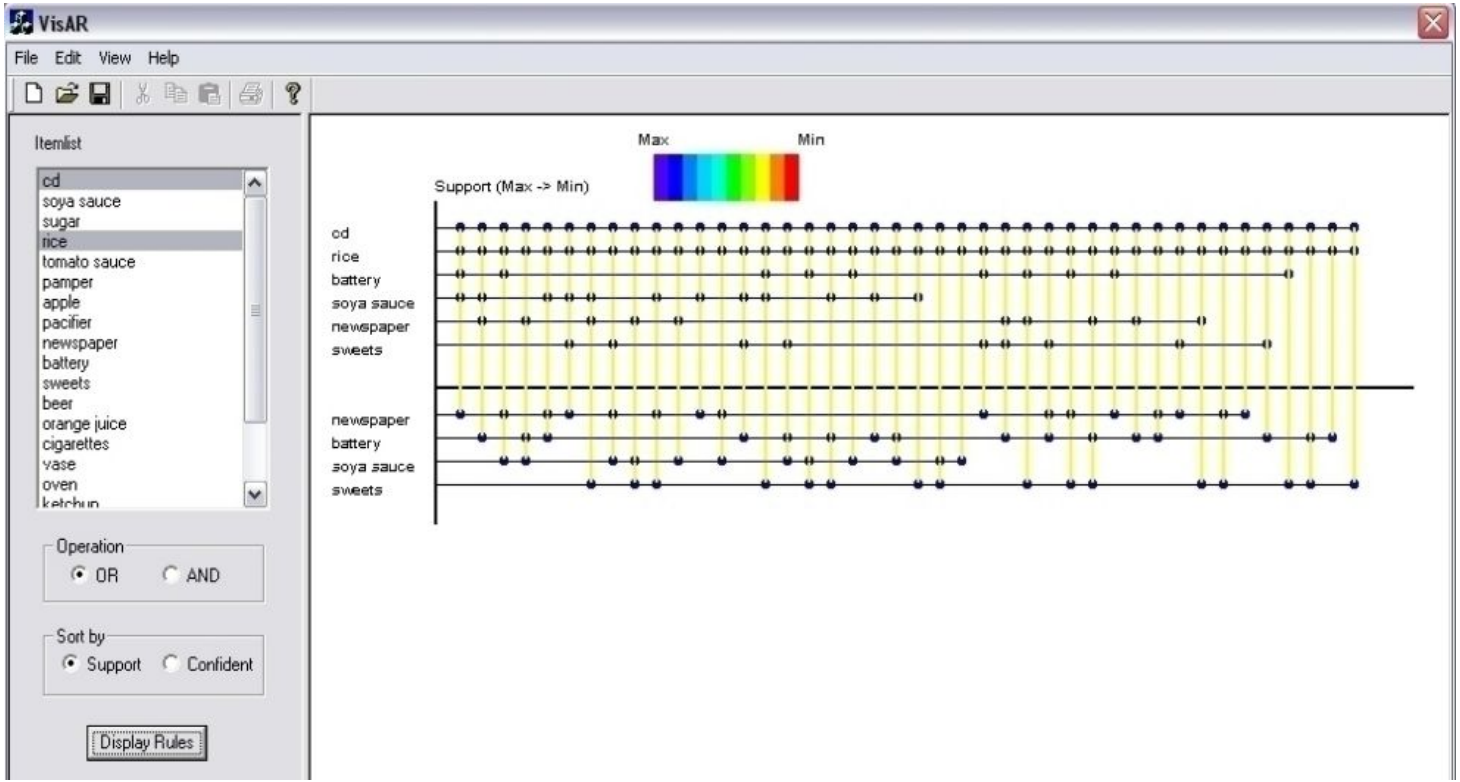


Figure 6: A visualization of item associations with support ≥ 0.4% and confidence ≥ 50%.

Wong, Pak Chung, Paul Whitney, and Jim Thomas. "Visualizing association rules for text mining." *Information Visualization, 1999.(Info Vis' 99) Proceedings. 1999 IEEE Symposium on*. IEEE, 1999.

VIsAR : A new Technique for visualizing Mined Association Rules

Combines graph based and matrix based technique and displays rules based on selected items in a more clean way



Techapichetvanich, Kesaraporn, and Amitava Datta. "VisAR: A new technique for visualizing mined association rules." *International Conference on Advanced Data Mining and Applications*. Springer Berlin Heidelberg, 2005.

RuleViz: A Model for visualizing knowledge discovery process

- An interactive model
- Classification Rules
- The viz tool CViz uses parallel coordinates to visualize rule where colors show the quality and accuracy of rules
- Filtering based on accuracy etc.

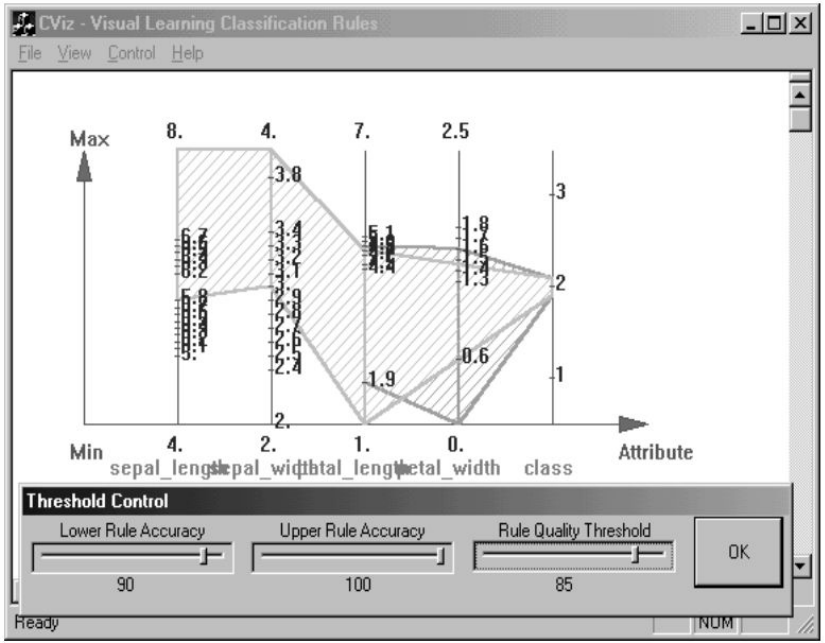


Figure 7: The rules for IRIS class 2 with quality greater than 85%

Han, Jianchao, and Nick Cercone. "RuleViz: a model for visualizing knowledge discovery process." *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2000.

3D Landscape-Rummaging

- Proposes a visualization for rule rummaging task where model depends on interactive rule focusing and on rule quality measures.
- Designing a rule network to allow user to navigate inside the visual representation.
- Rules are divided into subsets and combined through neighbourhood relations. (neighborhoods can be based on summarization, patterns, exception etc)
- After browsing a current subset, the user has choice to navigate to any neighboring subset.
- Only nodes interest the user and presented visually while the edges are only used for navigation.
- Two groups: specific and general
    - For example, in the subset named ABC where letters denote items, the specific rules are (A,B,C)->(D), (A,B,C)->(E), (A,B,C)->(F) and so on, and the general rules are (A,B)(C), (A,C)(B) and (B,C)(A).

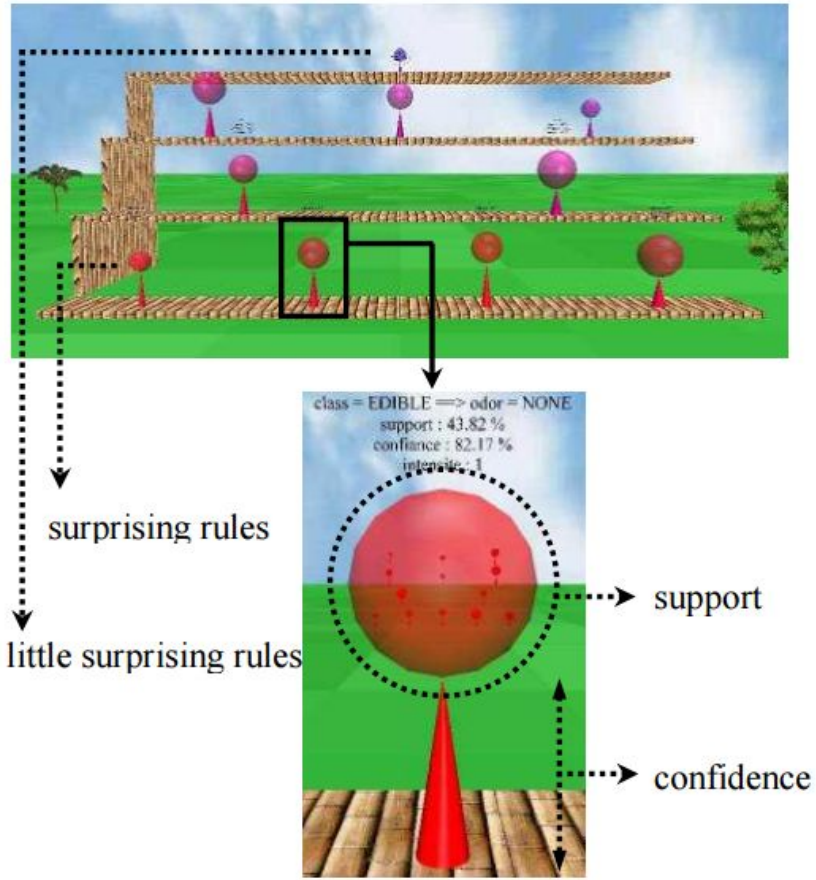


Figure 9. Metaphor description

Blanchard, Julien, Fabrice Guillet, and Henri Briand. "Exploratory visualization for association rule rummaging." *KDD-03 workshop on multimedia data mining (MDM-03)*. 2003.

Visualizing constraint-based temporal association rules

Interestingness Based on Three factors:

- Confidence
- Difference from neighbors
- J-Measure

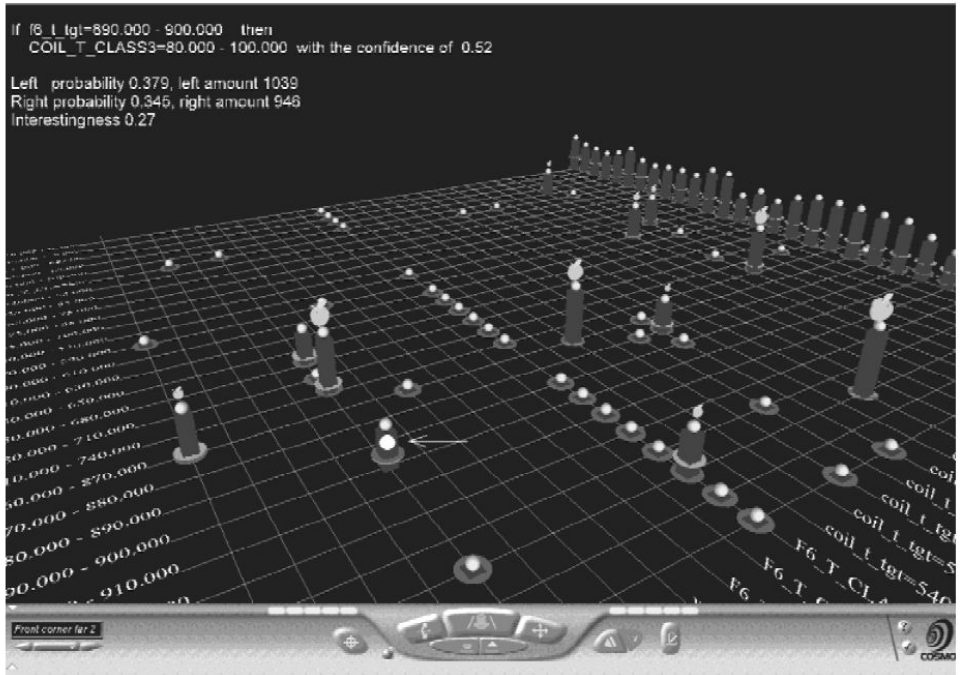
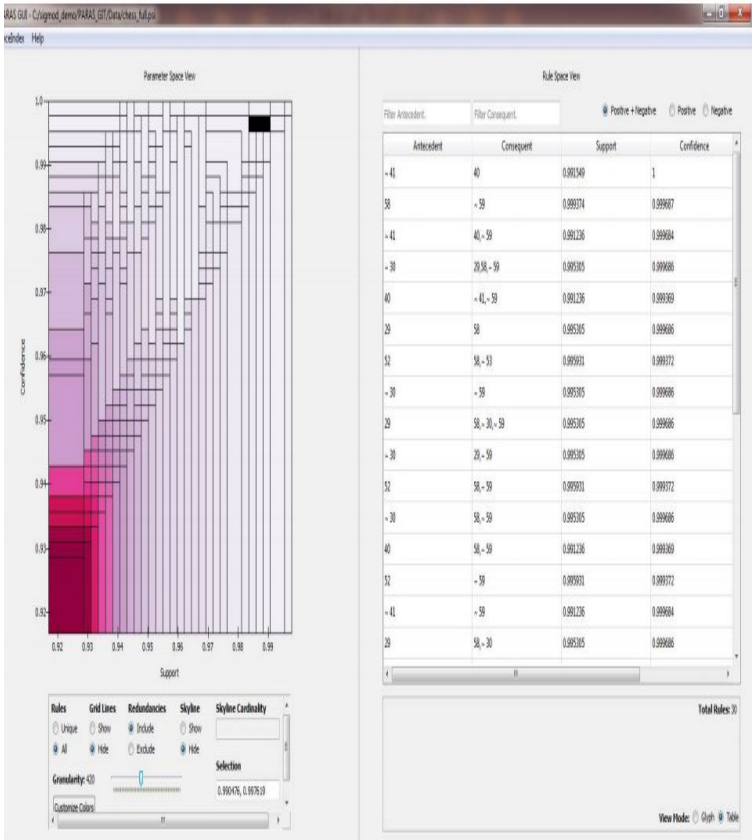
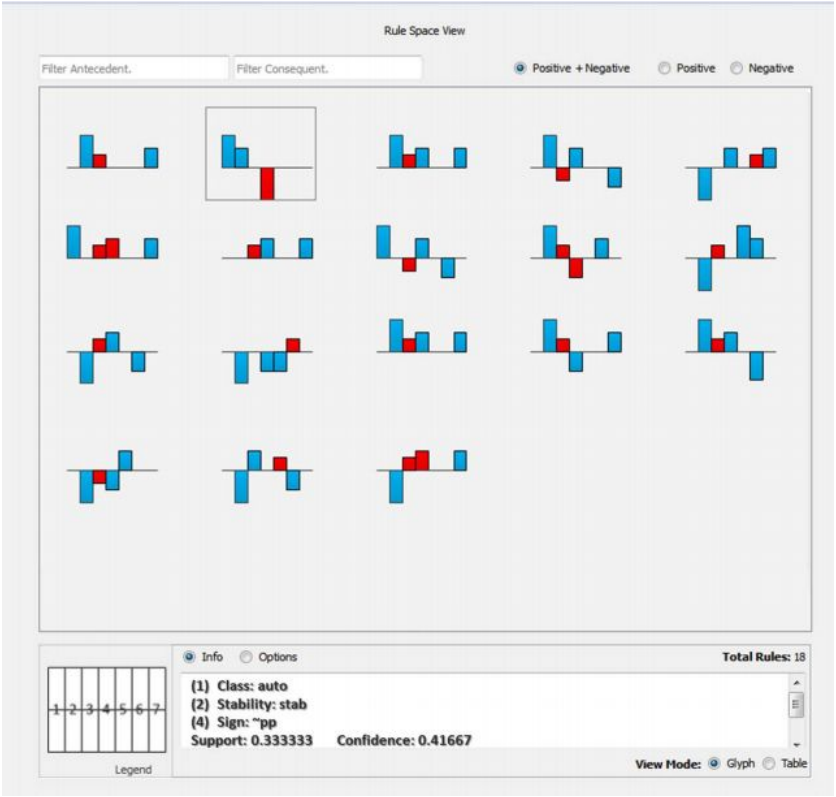


Fig. 2. The structure of the rule glyph

Korpipää, Panu. "Visualizing constraint-based temporal association rules."*AI EDAM* 15.05 (2001): 401-410.

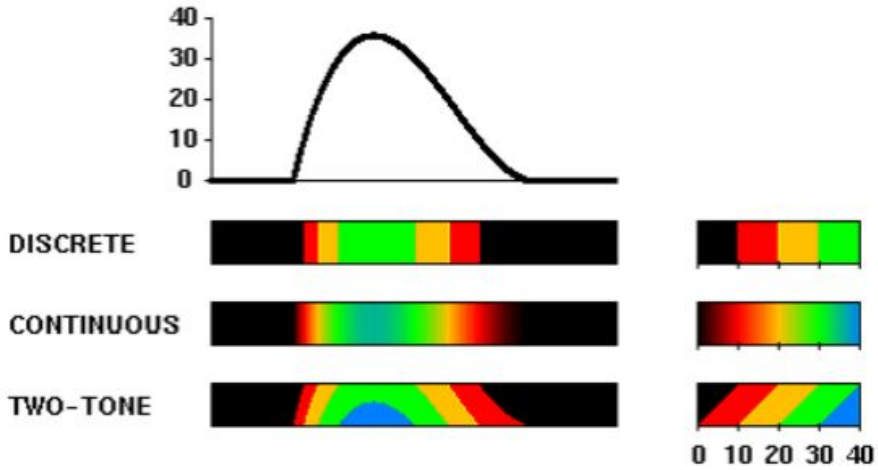
SPIRE: supporting parameter-driven interactive rule mining and exploration





Two-tone pseudo coloring: compact visualization for one-dimensional data

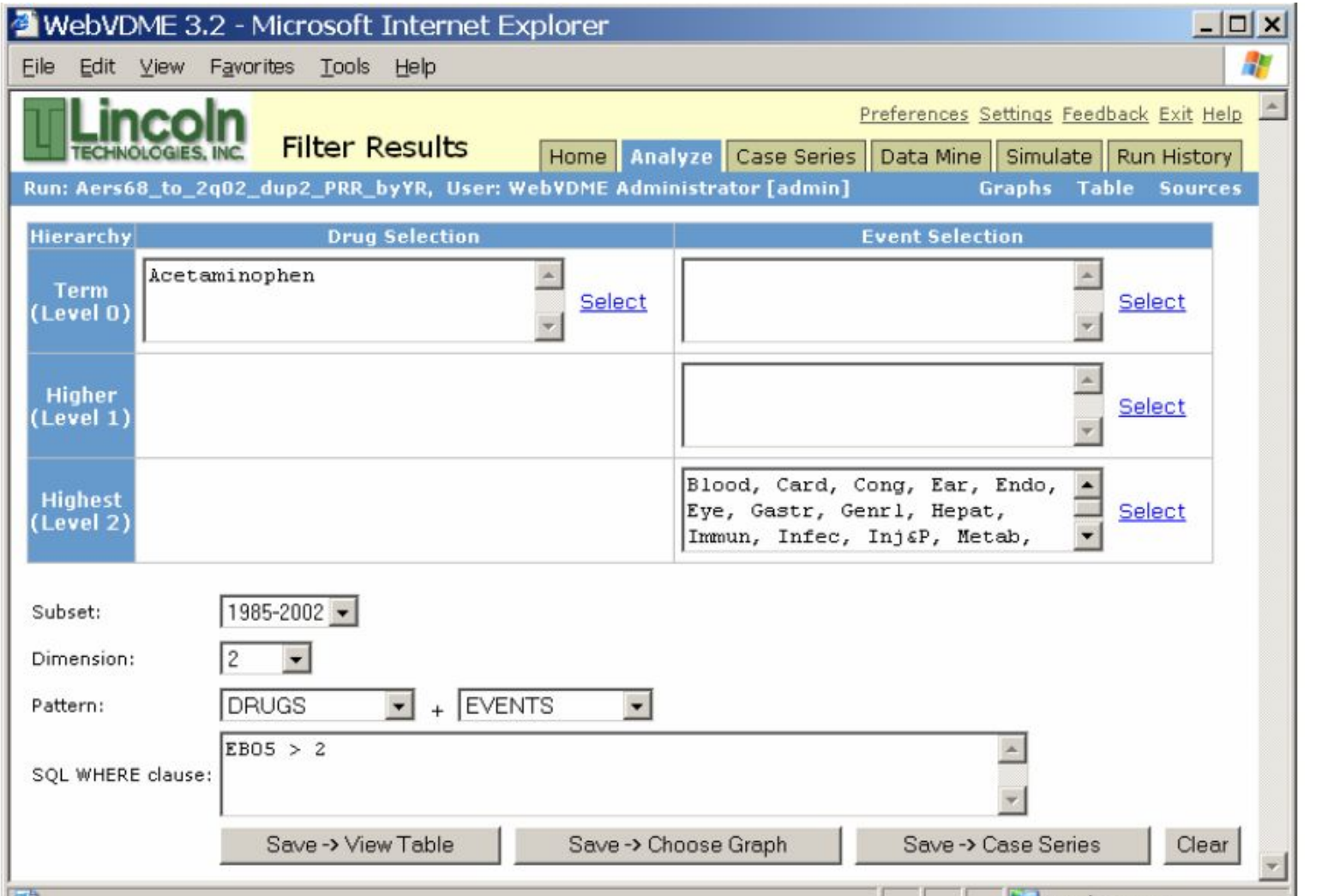
- Using pseudo coloring to visualize one dimensional but huge data set
- Each data point is represented by two colors.



Saito, Takafumi, et al. "Two-tone pseudo coloring: Compact visualization for one-dimensional data." *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005.*. IEEE, 2005.

Empirical Bayesian Data Mining for Discovering patterns in Post marketing Drug Safety --- SIGKDD'03

- Developed an analytical visual system to show drug ADR association
  - Using Multi-item Poisson Shrinkage and RR ratio
- Data Pre-processing
  - Mapping data to a unified view (as it was collected from different organizations)
  - Removing duplicates/ combining different versions - different case versions of same report
  - Using consistent drug names-- as there are trade names, generic names, active Ingredient
- The visual system includes;
  - Data mining tab --- to run the MPSG
  - Explore tab -- to filter for specific drugs etc and view the associated reports
  - Graphical view -- to show graphs for evolution of scores over time
  - Administrative actions -- for specific user rights etc etc


WebVDME 3.2 - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Lincoln TECHNOLOGIES, INC.
Filter Results
Preferences Settings Feedback Exit Help
Home Analyze Case Series Data Mine Simulate Run History
Run: Aers68_to_2q02_dup2_PRR_byYR, User: WebVDME Administrator [admin]
Graphs Table Sources
Hierarchy
Drug Selection
Event Selection
Term (Level 0)
Acetaminophen
Select
Select
Higher (Level 1)
Select
Highest (Level 2)
Blood, Card, Cong, Ear, Endo,
Eye, Gastr, Genrl, Hepat,
Immun, Infec, Inj&P, Metab,
Select
Subset:
1985-2002
Dimension:
2
Pattern:
DRUGS
+
EVENTS
SQL WHERE clause:
EBO5 > 2
Save -> View Table
Save -> Choose Graph
Save -> Case Series
Clear

CrystalClear: Active Visualization of association Rules

- Surveyed the state-of-the-art methods.
- Improved existing (Grid) methods by providing human interaction to the system and highlight important rules with multiple antecedents and consequents.
- The grid represents rules with confidence and support.
- Darker color means high confidence and support.
- Mouse over displays the whole rule.
- It also use icons to show if a new rule has been added, deleted or an interesting (high Support and Conf) has become uninteresting.
- It also provides a control panel to let user choose the colors to highlight rules.

ONG, Hian-Huat, et al. "Crystalclear: Active visualization of association rules." (2002).

Visualizing association rules in a framework for visual data mining

- Nodes can be rectangles, circles
- Red Antd, green consq
- Edge is the confidence, longer edge high confidence,
- Color saturation of edge shows support, light = low, dark = high
- A node is ant for one rule and consq for another, it's half red and half green
- Interaction through dragging is provided

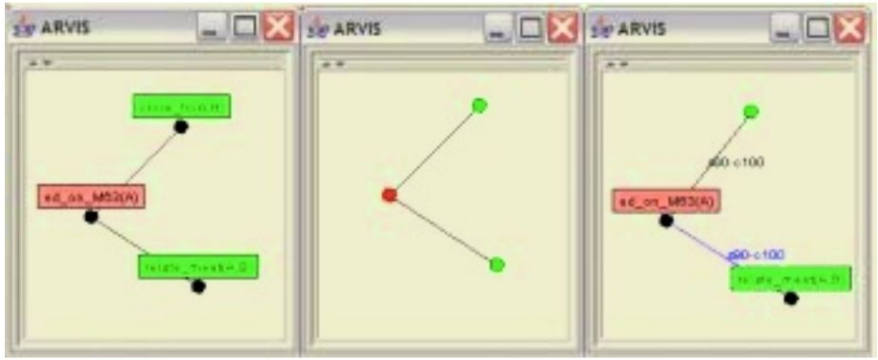


Fig. 2. Two association rules represented in a graph

- Default middle fig
- Rectangles give text so good when few rules
- Details visible over mouse over: most right

Buono, Paolo, and Maria Francesca Costabile. "Visualizing association rules in a framework for visual data mining." *From Integrated Publication and Information Systems to Information and Knowledge Environments*. Springer Berlin Heidelberg, 2005. 221-231.

Contd..

- Both sides show same rules
- Right fig: User has pulled a group of rules (with node half green half red) to corner to see how it is linked to the rest.
- The black color shows 100% confidence
- Now user can use Parallel co-ordinates to see this group in detail and prune some uninteresting rules

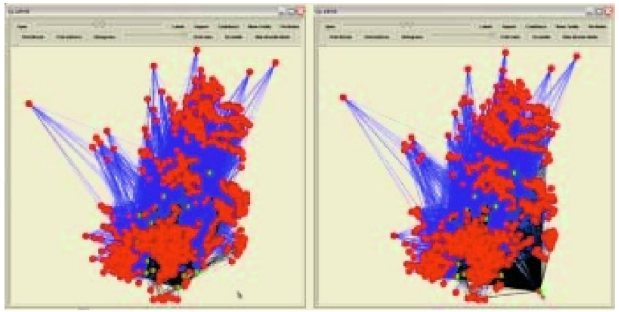


Fig. 3. 9785 association rules shown in one screen

Contd..

- User can identify interesting rules using the Item Utility or IU (how important an item is for a rule
- The IU indicates how good is an item into a rule
- Three cases
    - IU>0 means useful, IU<0 means dangerous and IU=0 means item is neutral for rule property.
- Items are represented on x axis and rules are represented by line.
- The parallel coordinates show the IU value normalized between -1
- each line crosses the axis depending upon the IU value.
- Can help to prune rules with less IU threshold
- Graph visualization is useful to describe the overview of the rules and the relationships among the items, parallel coordinates work better if used as a visual pruning tool.
- The pruned rules can again visualized in graph view for further analysis
- The framework is open to integrate other visualization tools too.

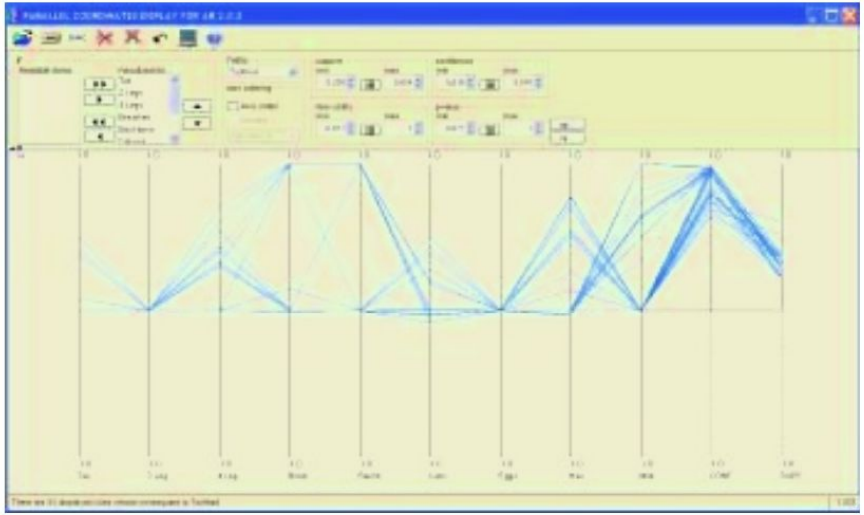


Fig. 4. Association rules visualization using parallel coordinates

Buono, Paolo, and Maria Francesca Costabile. "Visualizing association rules in a framework for visual data mining." *From Integrated Publication and Information Systems to Information and Knowledge Environments*. Springer Berlin Heidelberg, 2005. 221-231.

Points related to Our Research

- We can use the idea of presenting many drug interactions as a scatter plot or graph, where user can interact with each interaction separately and then give details

Pruning and visualizing generalized association rules in parallel coordinates

- An approach for visualizing frequent itemsets (FIS) and many-to-many association rules by a novel use of parallel coordinates
- A frequent itemset or an association rule is visualized by connecting items, one on each parallel coordinate. No of coordinates depends on the number of items in the longest FIS
- For generalized association rules where items are organized into item taxonomy, each coordinate can be used to visualize an expendable item taxonomy tree whose leaf nodes are items and whose non leaf nodes are item categories.
- Only the frequent itemsets at the border of frequent itemsets in the itemset lattice. Subsets of the visualized frequent itemsets are implied and are not displayed. F

Pruning and visualizing generalized association rules in parallel coordinates

- 4a shows 3 itemsets
  - Fg, cdb, adbe
- 4b: Rule ab -> cd
  - ab->c, ab->d are also closed so are not visualized
- If two or more itemsets or rules have items in common, i.e. adbe and cdb in Fig. 4a, used polynomial curves instead of polylines to distinguish one from the rest
- Fig 5: two rules ab-> ce and db->ce
- Also used it to show generalized rules by visualizing item taxonomies

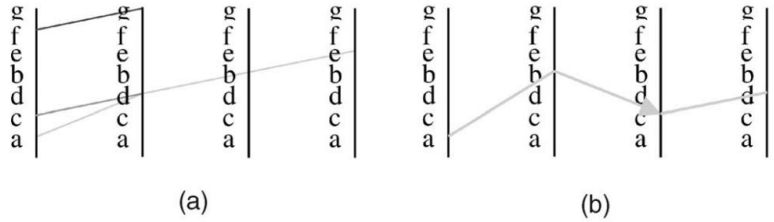


Fig. 4. Visualizing (a) frequent itemsets and (b) an association rule.

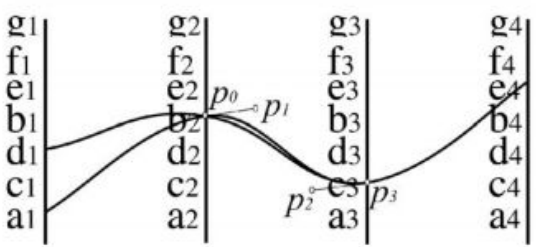
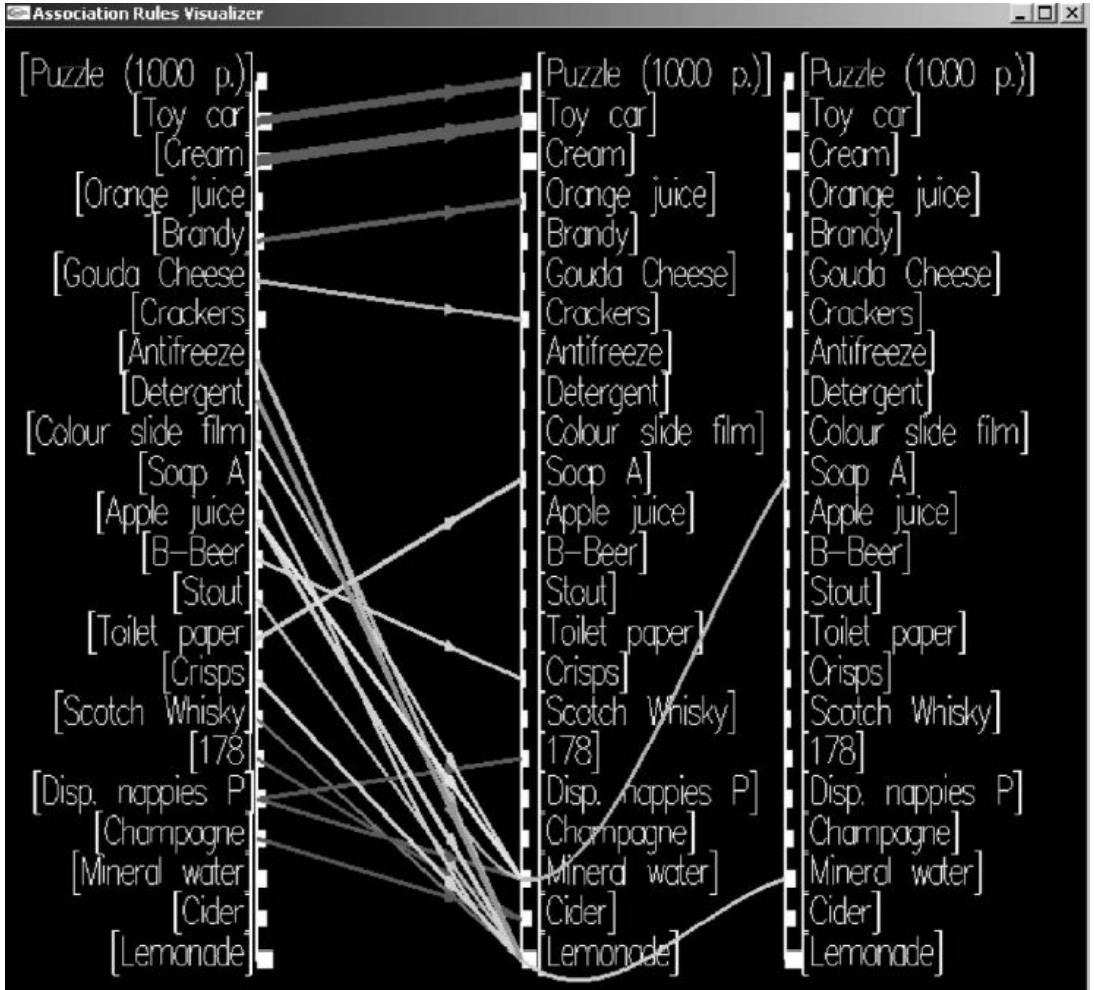


Fig. 5. Visualizing association rules using Bézier curves.

Parallel Coordinates

- Bezier Curves:
- Direction of arrow shows Ant -> Cons relationship
  - LHS is ant
  - RHS (2 columns) is cons
- Items are grouped and sorted by support and shown by width of curve, support of item is shown by small bar
- Conf is shown by color of the curve
- Each rule is selectable and displays the implied rules on side



Yang, Li. "Pruning and visualizing generalized association rules in parallel coordinates." *IEEE Transactions on Knowledge and Data Engineering* 17.1 (2005): 60-70.

Visualizing Association Rules Using Linked Matrix, Graph, and Detail Views

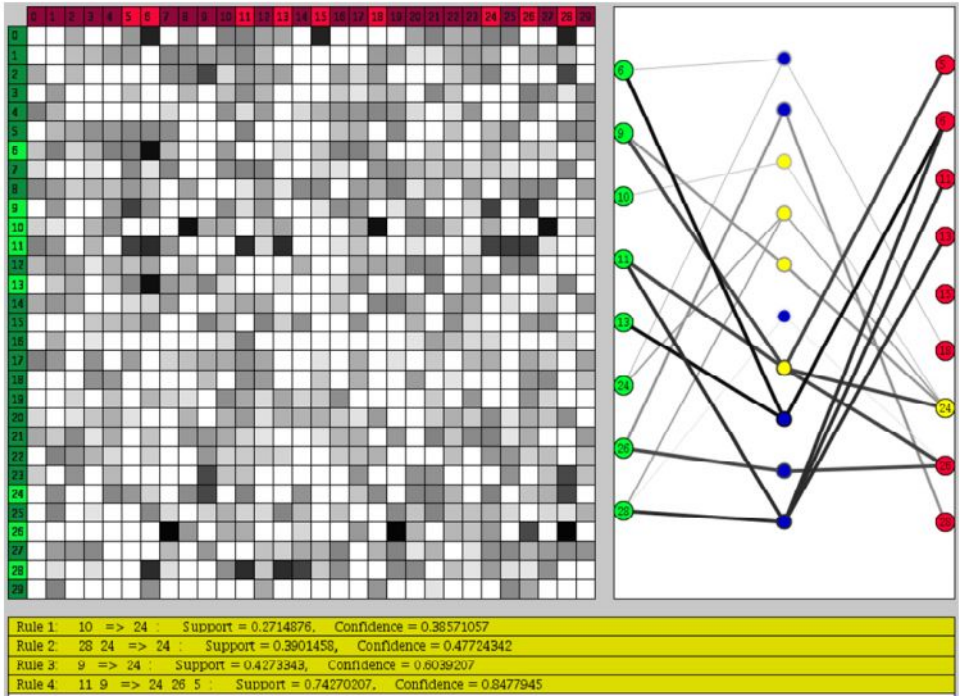
Matrix View

- Rows - LHS
- Columns - RHS
- Grey scale darkness shows support
- Each cell represents a rule with must items(r,c) and can have other items too
- Light green color are the selected items that are LHS for a rule having a selected cell (RHS)

Graph View:

- Shows the subset selected from matrix view
- Color of line shows support
- Thickness shows Confidence
- Blue circle shows the rule between LHS RHS.
- Yellow are the highlighted rules (all rules that have same RHS)

Evaluated via a user study, to see if users can find interesting patterns with var No of rules upto 500..



Y. Sekhavat and O. Hoeber, "Visualizing Association Rules Using Linked Matrix,Graph, and Detail Views,"*International Journal of Intelligence Science*, Vol. 3 No. 1A, 2013, pp. 34-49. doi: 10.4236/ijis.2013.31A005.

Key Points:

- Matrix view not good when have huge itemset
- Can borrow the idea of matrix showing one drug and ADRs but may contain others for 100 interactions or so. (might not be good to show all interactions)
- Idea of detailed view is same as ours

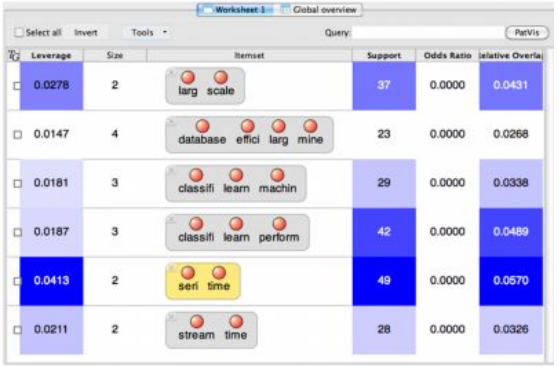
MIME: A Framework for Interactive Visual Pattern Mining

- Uses many interestingness measures (does not specify which)

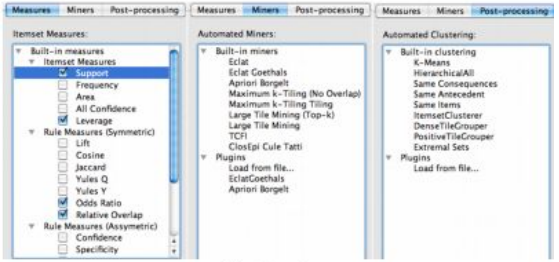
Three main Components

- Source dock has the items, sorted by rank
  - Items are sorted based on their rank
  - Items can be grouped here
- Work dock gives rules-> color and numbers are used to show interestingness
  - Items can be draged to work dock to show the rules
  - A user can see other items that are potentially interesting for an existing itemset.
- Tool box shows built in and plugin measures, mining algorithms and post-processing Global overview: to compare different work spaces

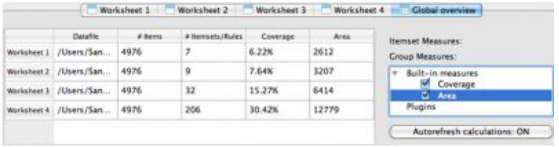
(a) Source dock



(b) Work dock



(c) Toolbox



(d) Global overview

Goethals, Bart, Sandy Moens, and Jilles Vreeken. "MIME: a framework for interactive visual pattern mining." *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2011.

Key points

- Our tool can have these components too
- Our source dock can be drugs with their pictures

Glyph Sorting:

- A Glyph based conceptual framework as part of a visualization process for interactive sorting of the multivariate data.
- Design principles for for developing visually sortable glyphs
- Combined benefits of sorting with glyph-based viz
- https://www.youtube.com/watch?v=doDMaAlgzII

Design Principles for sorting glyphs

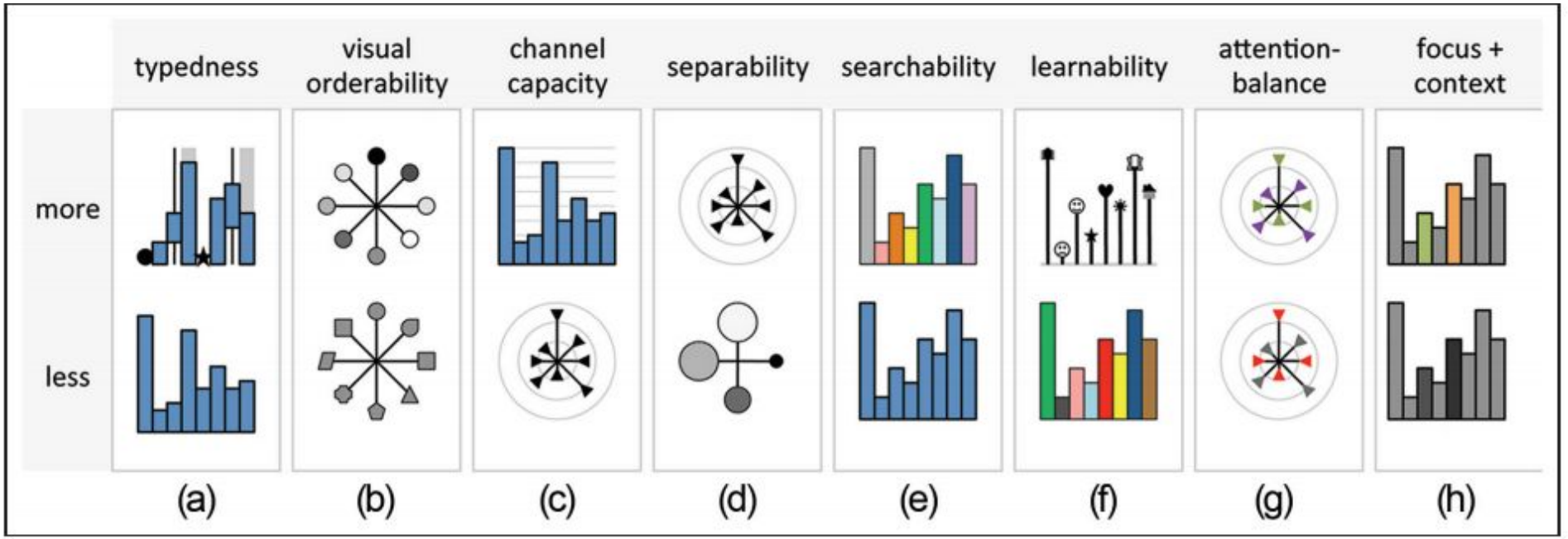


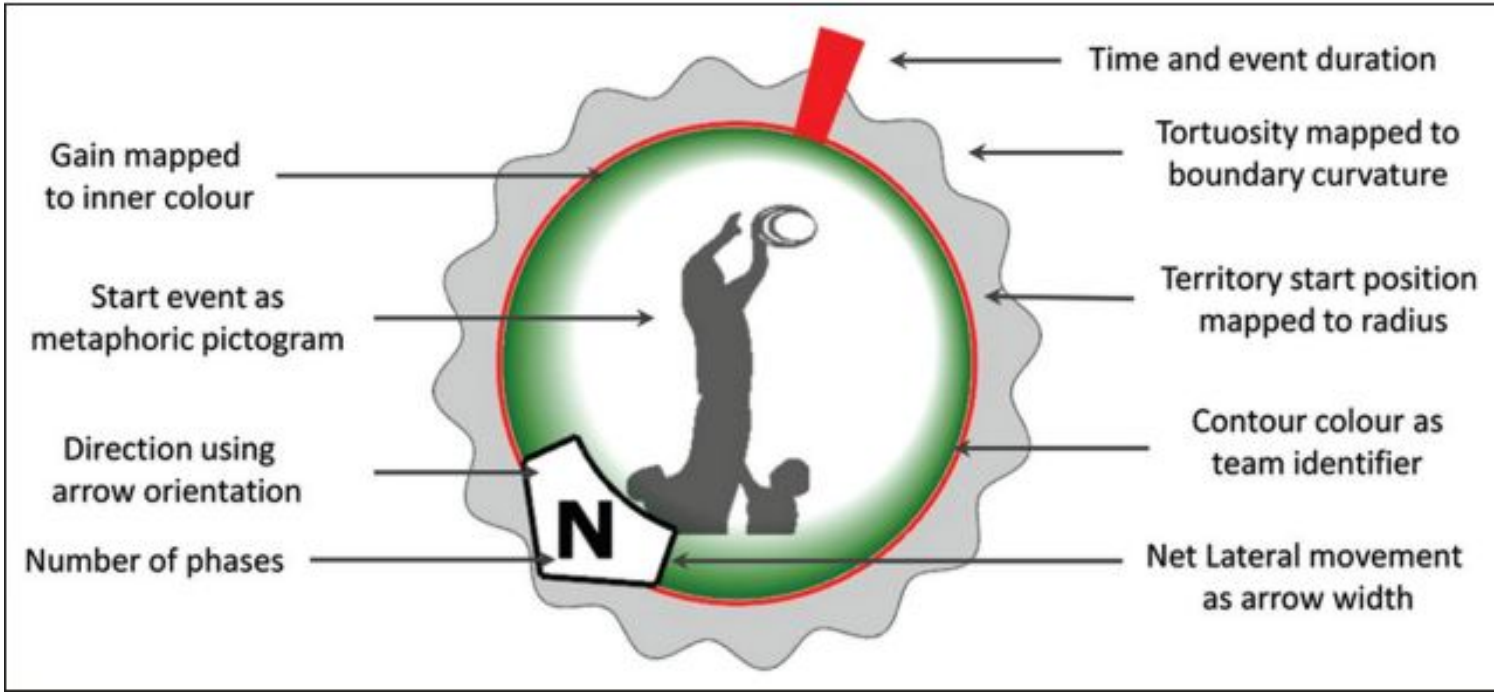
Figure 2. (a)–(h) Variations of glyph design in accordance to the design principles of sortable glyph. For each principle, the top row depicts a glyph with greater emphasis and the bottom row depicts a glyph with less emphasis.

Explanation

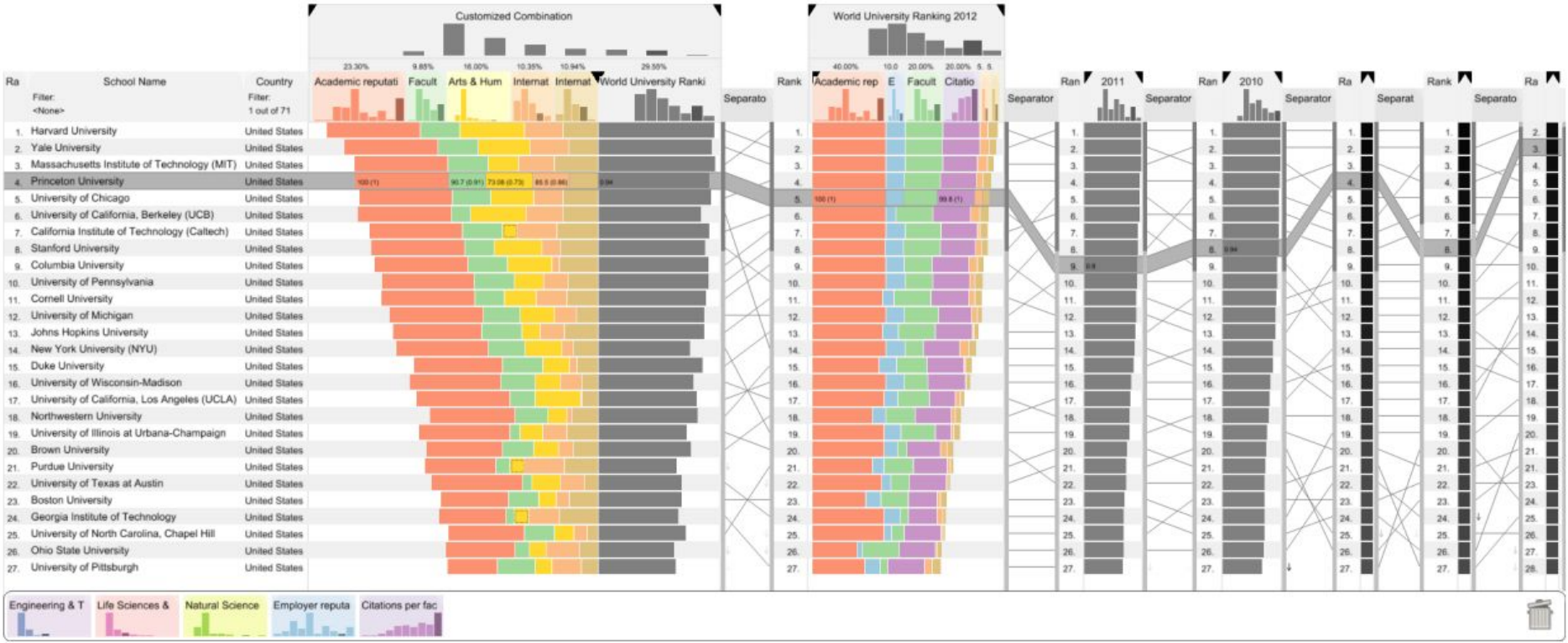
- Typedness: 1st & 4th are nominal, and 3rd & 7th are interval
- Orderability: ordering easy in top one, metaphor pictograms not good
- Channel Cap: number of values encoded by visual channels (bars have size, color and length while radial cannot accomodate size)
- Separability; Two separate two channels (intensity and size)
- Searchability: easy to search
- Learnability: pictograms good to associate the viz with data
- Attention balance: The red color might distract the user
- Focus context: highlights on the attributes user is interested in

Glyph Design

- Each glyph rich in information
-



LineUp: Visual Analysis of Multi-Attribute Rankings



Gratzl, Samuel, et al. "Lineup: Visual analysis of multi-attribute rankings."*IEEE transactions on visualization and computer graphics* 19.12 (2013): 2277-2286.

LineUp: Visual Analysis of Multi-Attribute Rankings

- An interactive technique designed to create, visualize, and explore rankings of items based on a set of heterogeneous attributes.
- LineUp can be used to create and compare multiple rankings of the same set of items
- Combined attributes
  - Serial: Weighted sum individual attribute scores.
  - Parallel: Maximum of set of attributes
- In case of rank change (based on some filter etc.), color encoding shows if rank went higher(green) or lower (red)
- Comparison of ranking using slope lines
- Collapsing columns to heatmap when needed
- A memo pad to show the attributes that are not currently of interest
- user study with 8 participants with 12 tasks

Gratzl, Samuel, et al. "Lineup: Visual analysis of multi-attribute rankings."*IEEE transactions on visualization and computer graphics* 19.12 (2013): 2277-2286.

Rank As You Go: User-Driven Exploration of Search Results. *IUI'16*

- URANK to investigate interactive methods for understanding, refining and organizing documents on-the-fly
- uRank includes views summarizing the contents of a recommendation set and interactive methods conveying the role of users' interests through a recommendation ranking.

di Sciascio, Cecilia, Vedran Sabol, and Eduardo E. Veas. "Rank As You Go: User-Driven Exploration of Search Results." *Proceedings of the 21st International Conference on Intelligent User Interfaces*. ACM, 2016.

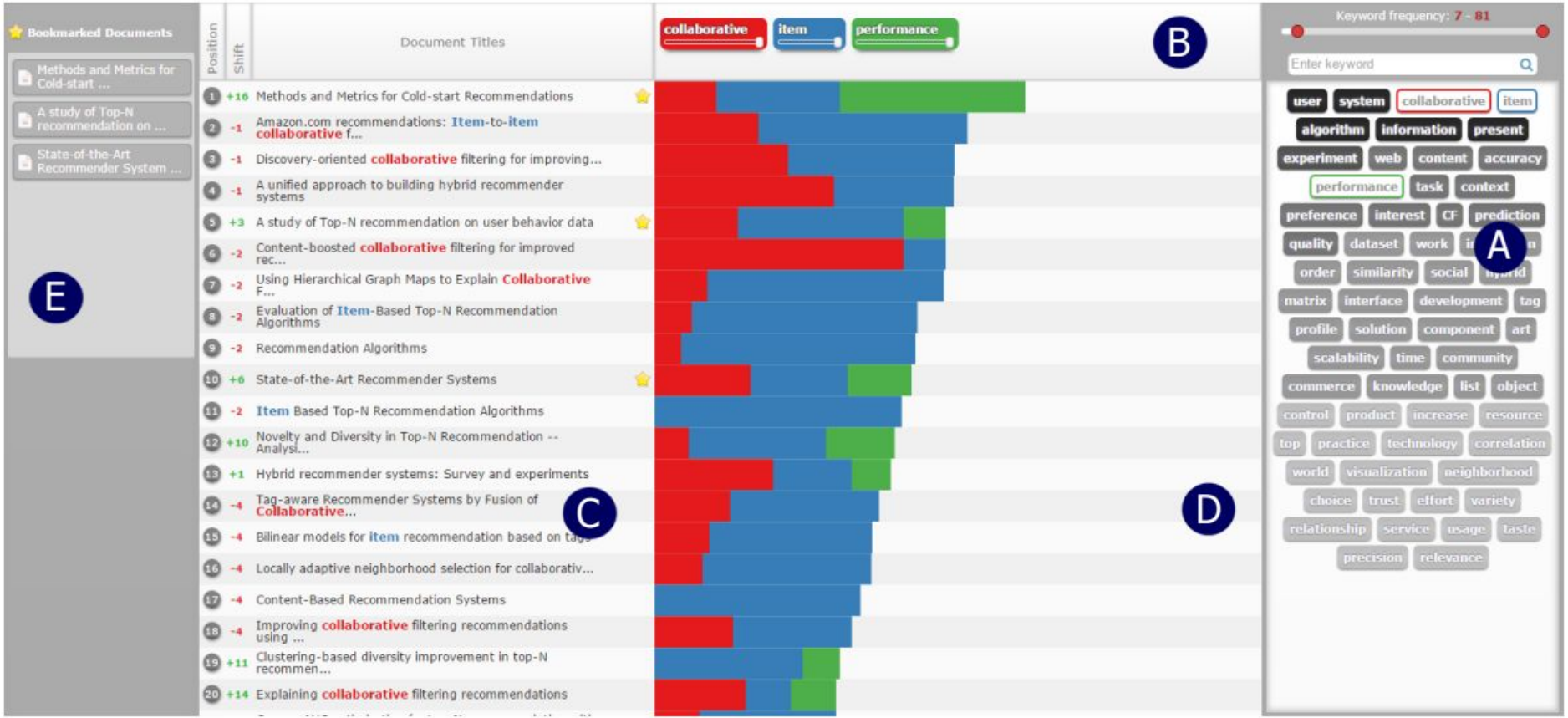


Figure 2. ***uRank*** **User Interface displaying documents related to** ***recommender systems*****, with ranking updated to match the keywords "collaborative", "item" and "performance". A. The** ***Tag Box*** **presents a keyword-based summary of the document collection, B. the** ***Query Box*** **contains keywords selected by the user, C. the** ***Document List*** **and D. the** ***Ranking View*** **present a list with augmented document titles and stacked bars indicating relevance scores, and E. the** ***Bookmark Overview*** **shows bookmarked documents.**

More Reading

PhD Dissertation

- Glyph Design for Temporal and Multi-Dimensional Data: Design Considerations and Evaluation
- Fuchs, Johannes Hermann. *Glyph Design for Temporal and Multi-Dimensional Data: Design Considerations and Evaluation*. Diss. 2015.

Papers related to Drug-ADRs Visualizations

Rainbow boxes

- To show properties (ADRs and Contraindications) of same class of drugs (similar drugs)
- Works poor for different drugs
- Used VCM (visualization of concepts in medicine) , icons in the graph
- For CI-Red for current state, orange future state, brown past state
- For ADRs- red for severe and saturation for frequency
- Conducted user study with physicians by comparing with tables to compare speed and accuracy
- Ordering algorithm to reduce the holes in the viz.

Lamy, Jean-Baptiste, Hélène Berthelot, and Madeleine Favre. "Rainbow boxes: a technique for visualizing overlapping sets and an application to the comparison of drugs properties." *International Conference Information Visualisation 2016*. 2016.

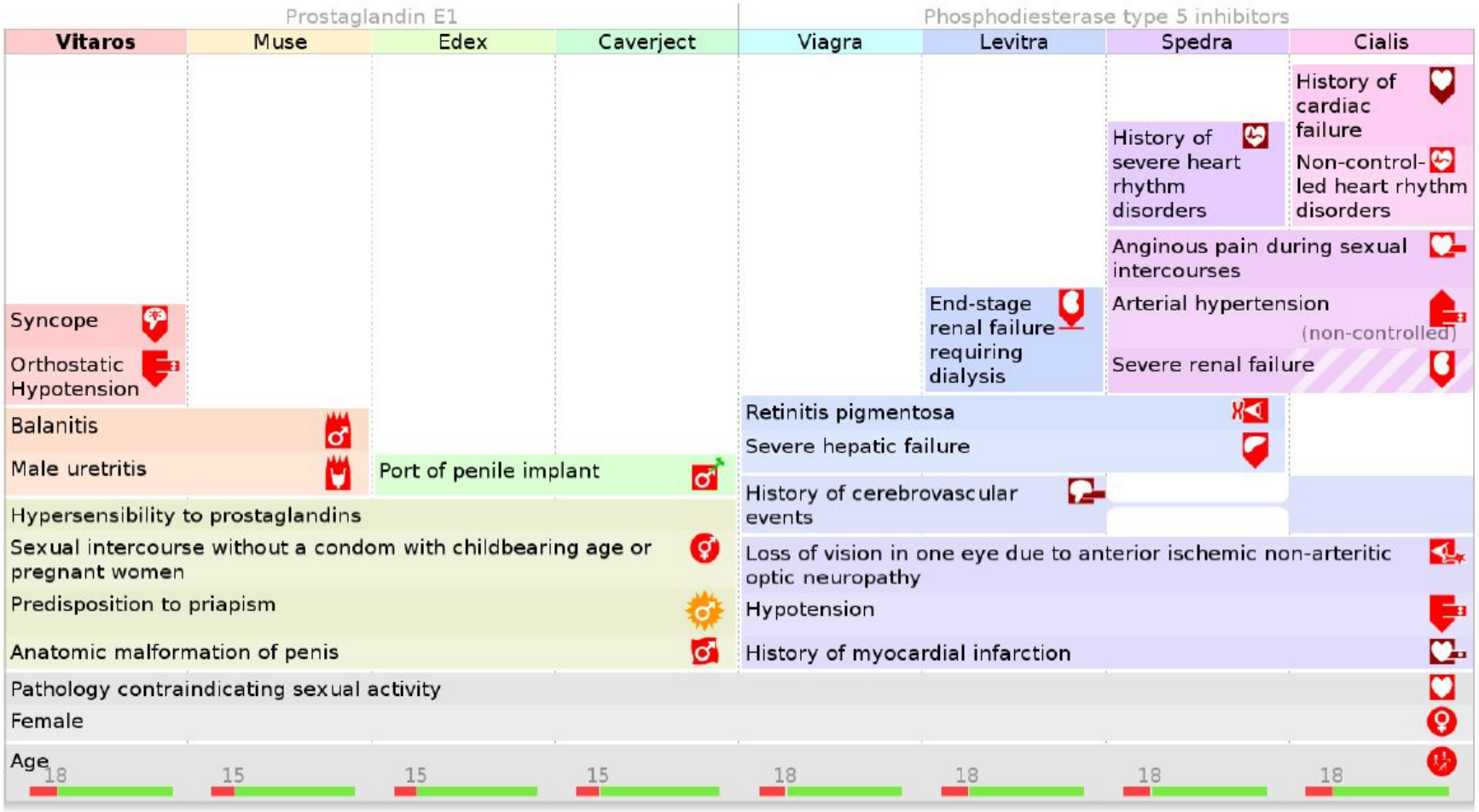
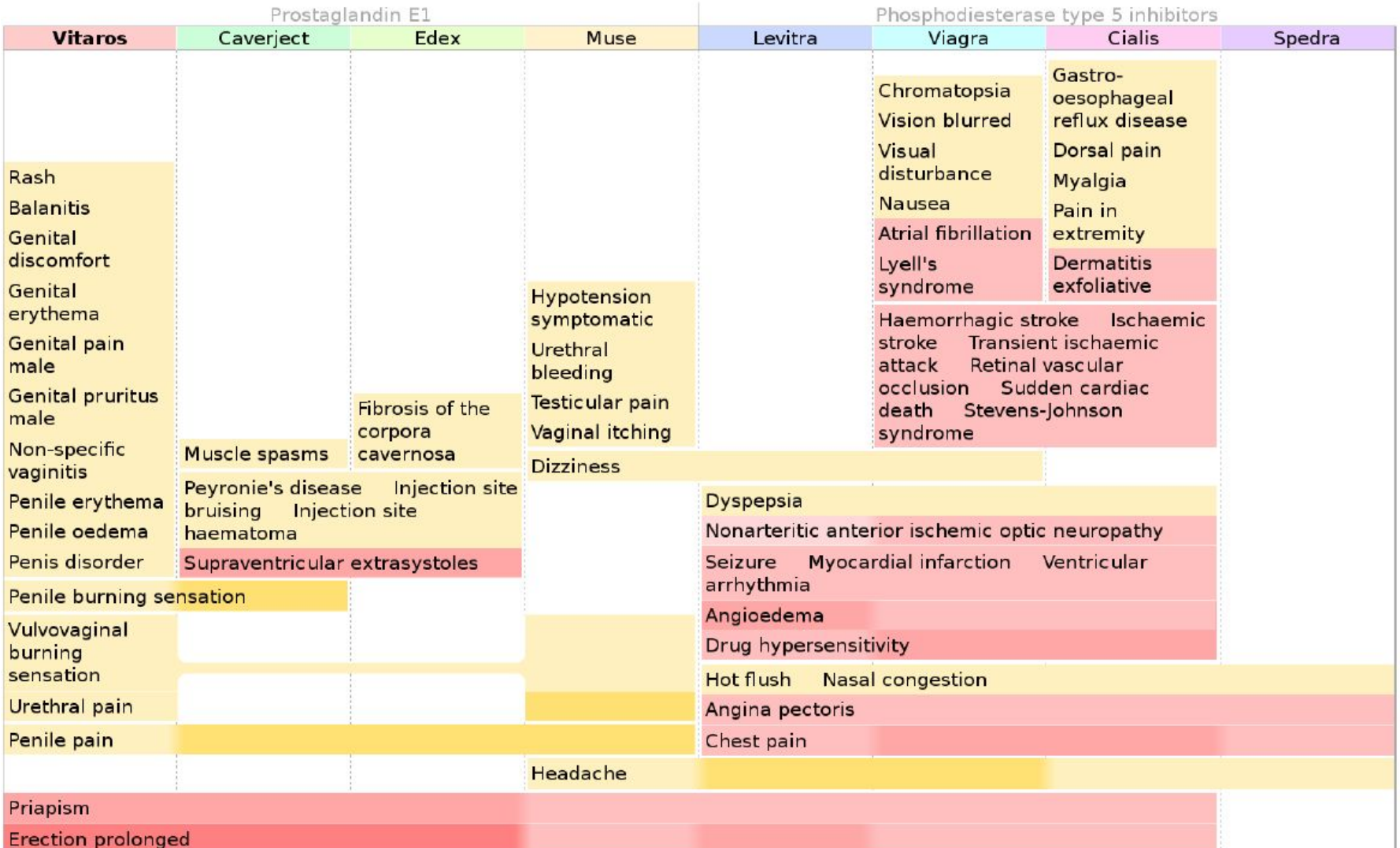


Figure 1. Rainbow boxes displaying the 79 contraindications (26 distinct) of 8 drugs for erectile dysfunction. The drugs are shown in columns and the contraindications in rectangular horizontal boxes (possibly with holes) covering the columns corresponding to the drugs sharing the contraindications.


Prostaglandin E1
Phosphodiesterase type 5 inhibitors
Vitaros
Caverject
Edex
Muse
Levitra
Viagra
Cialis
Spedra
Chromatopsia
Vision blurred
Visual disturbance
Nausea
Atrial fibrillation
Lyell's syndrome
Gastro-oesophageal reflux disease
Dorsal pain
Myalgia
Pain in extremity
Dermatitis exfoliative
Rash
Balanitis
Genital discomfort
Genital erythema
Genital pain male
Genital pruritus male
Non-specific vaginitis
Penile erythema
Penile oedema
Penis disorder
Penile burning sensation
Vulvovaginal burning sensation
Urethral pain
Penile pain
Hypotension symptomatic
Urethral bleeding
Testicular pain
Vaginal itching
Fibrosis of the corpora cavernosa
Muscle spasms
Peyronie's disease
Injection site bruising
Injection site haematoma
Supraventricular extrasystoles
Haemorrhagic stroke
Ischaemic stroke
Transient ischaemic attack
Retinal vascular occlusion
Sudden cardiac death
Stevens-Johnson syndrome
Dizziness
Dyspepsia
Nonarteritic anterior ischemic optic neuropathy
Seizure
Myocardial infarction
Ventricular arrhythmia
Angioedema
Drug hypersensitivity
Hot flush
Nasal congestion
Angina pectoris
Chest pain
Headache
Priapism
Erection prolonged

Rxplore:

- System to visualize ADRs of multiple drugs simultaneously
- A Decision support system for patients with multiple medications
- 250 commonly used drugs, ADR information extracted from drug labels and mapped to MedDRA terms
- To standardized extracted frequencies (occurred in 12% of patients, occurs frequently, was reported) , Rxplore score is devised
- Conducted user-study to compare *Rxplore* with *UpToDate* for speed and accuracy of answering the question
- System is for Known ADRs and not interactions
- What if drugs have not-common ADRs and they are many in numbers?
- http://www.uptodate.com/home/help-demo

Duke, Jon D., Xiaochun Li, and Shaun J. Grannis. "Data visualization speeds review of potential adverse drug events in patients on multiple medications." *Journal of biomedical informatics* 43.2 (2010): 326-331.

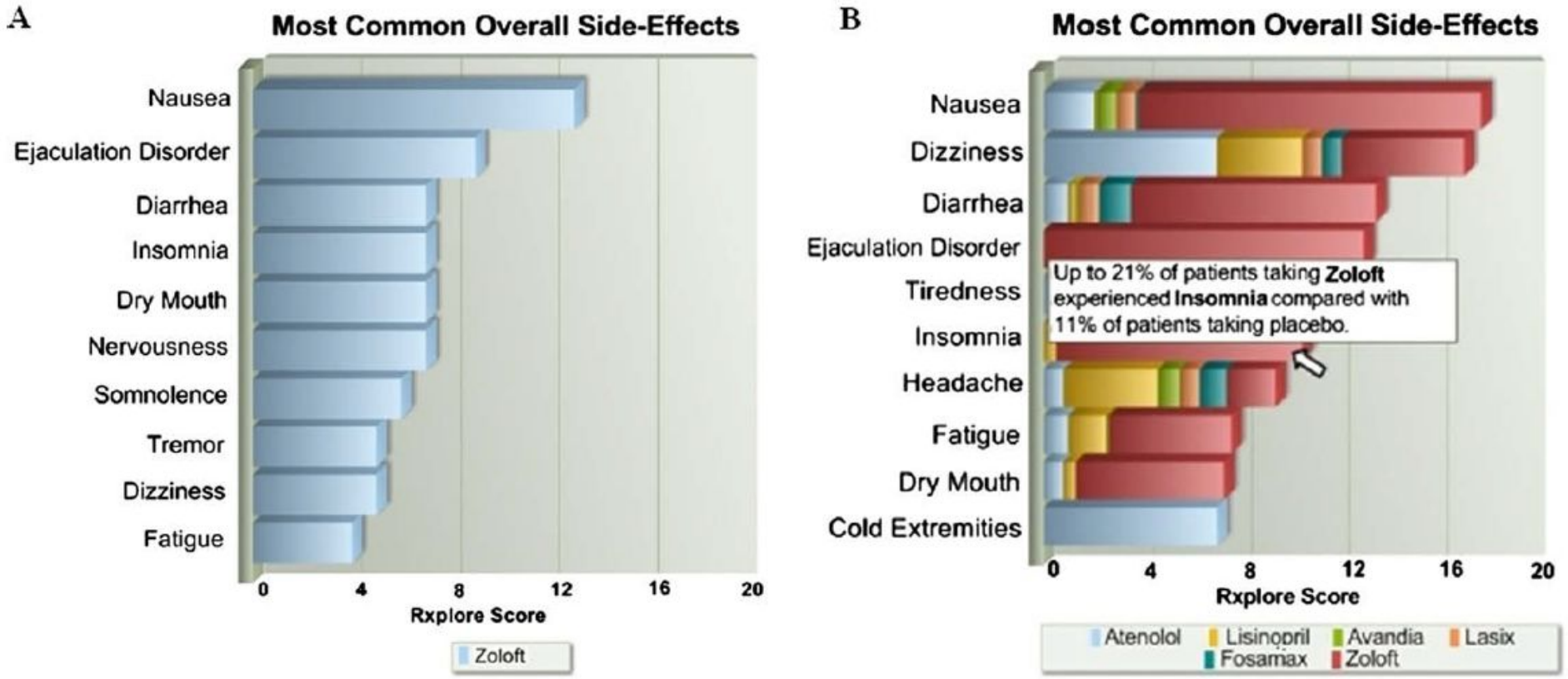


Fig. 1. Results display in Rxplore. (A) The most common reactions seen with Zoloft alone. (B) The most common "combined" effects of a six drug regimen.

A novel visualization for evaluating medication SE in Multi drug Regimens

- Same Rxplore
- Duke, Jon, Anthony Faiola, and Hadi Kharrazi. "A novel visualization tool for evaluating medication side-effects in multi-drug regimens." *International Conference on Human-Computer Interaction*. Springer Berlin Heidelberg, 2009.

ADRViz: an Information Visualization Platform for Adverse Drug Reactions.

Jian-Xiang, Wei, et al. "ADRVis: an Information Visualization Platform for Adverse Drug Reactions." *International Journal of u-and e-Service, Science and Technology* 8.10 (2015): 139-150.

Table 1. Instructions of Three Legends

Legend	Type	Note
	CASE I	Solid circle, indicating that the ADR of one specific drug was stated in drug instructions and also actually detected.
	CASE II	Solid rectangle, indicating that the ADR of one specific drug was actually detected but not stated in drug instructions.
	CASE III	Hollow dashed circle, indicating that the ADR of one specific drug was stated in drug instructions but not actually detected.

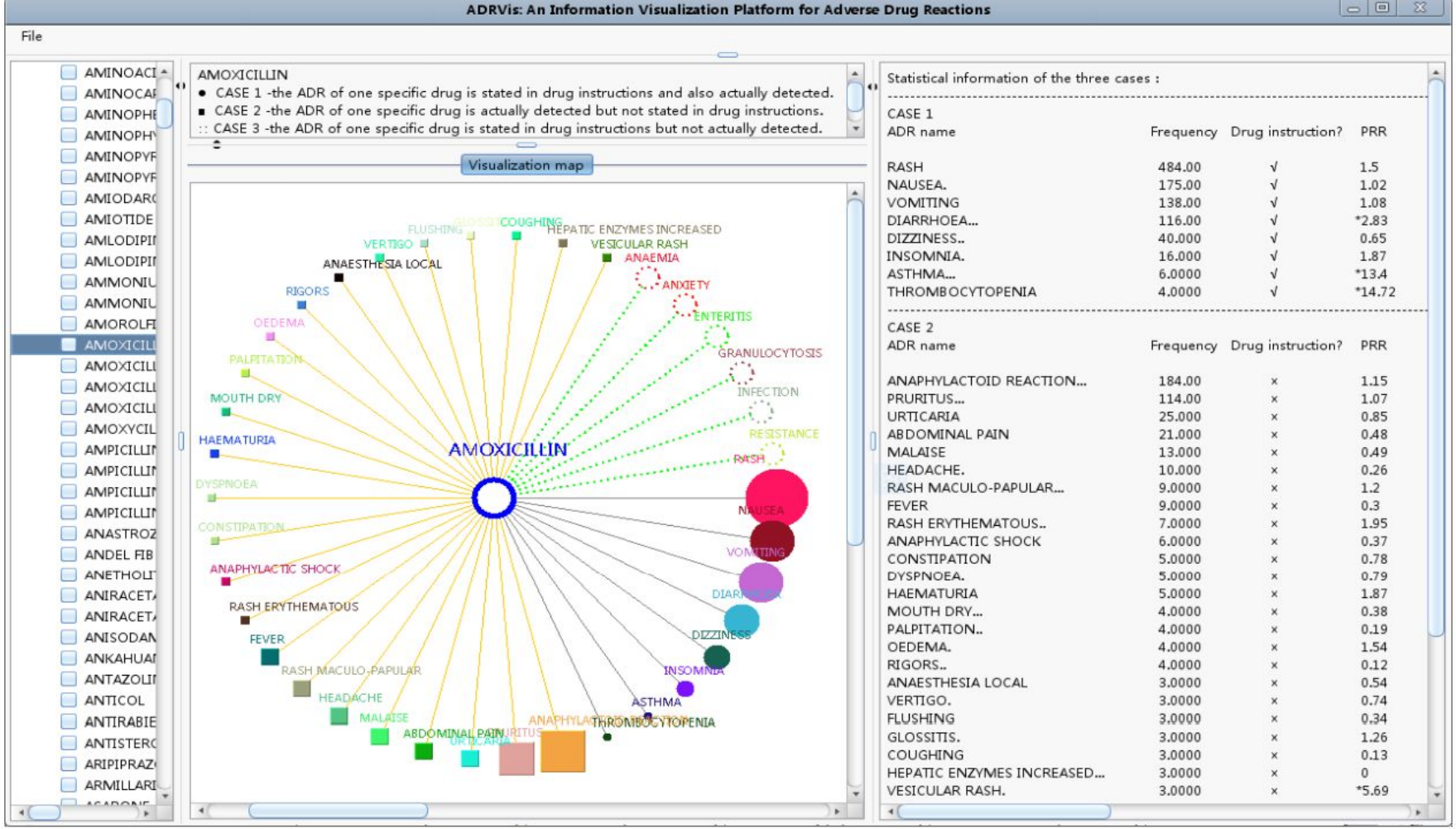


Figure 5. Visualization Map of Amoxicillin's ADRs

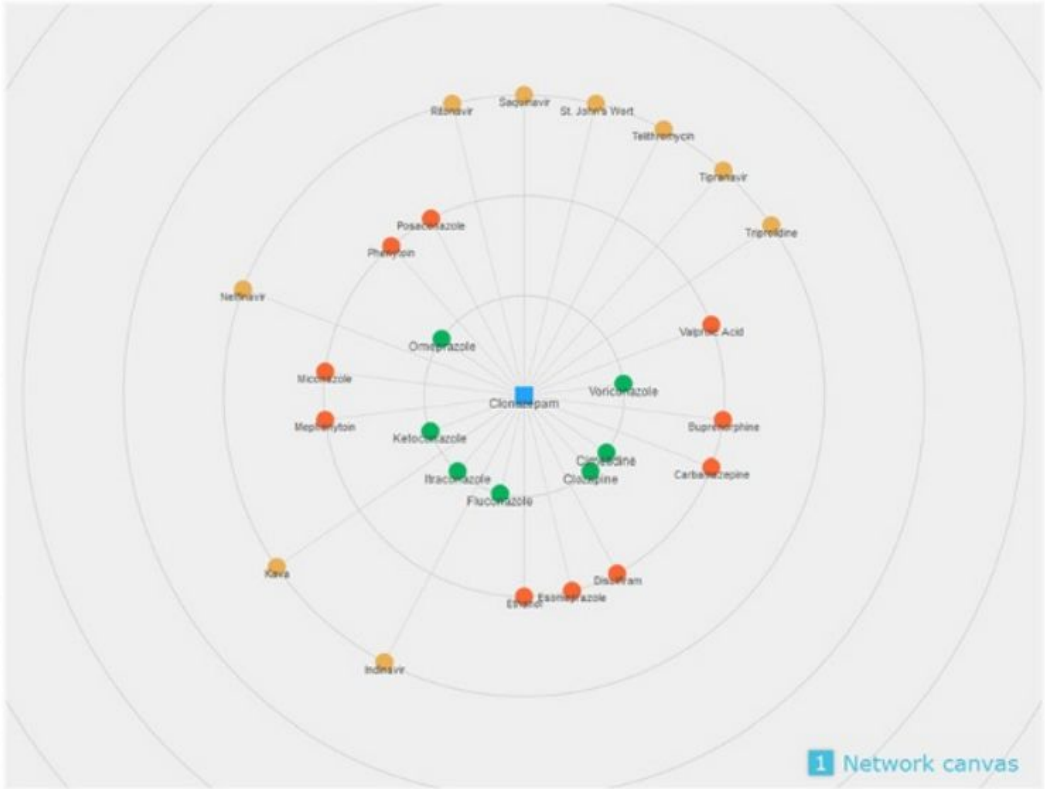
GraphSAW:

- Data from various sources SIDER, ABDA, Drugbank
- Mapped to common MedDRA ADRs
- Correlative: Drugs/ADRs that are found in both databases
- Gives protein and molecular level details.
- Other systems that shows DDI or drug related ADRs:
  - http://stitch.embl.de/
  - http://bioinformatics.charite.de/promiscuous/
  - http://www.genome.jp/kegg/medicus.html
  -

GraphSAW

3 Navigation bar

- Graph settings (b)
  - **Graph key**
    - = Drug
    - = Interactions from ABDA
    - = Interactions from DrugBank
    - = Correlative interactions
  - Edge width:
  - Node size:
- Molecular medication analysis
- Single drug side effects
- Cumulative drug side effects
- Single drug interactions (a)
  - **Drug name**
  - Clonazepam
  - Search
- Combined drug interactions
- Drug-Molecule interactions
- Drug-Pathway mapping
- Pathway-Drug mapping
- Pathway-Disease network



1 Network canvas

2 Side bar

Interaction partners for Clonazepam [26]

Interactions obtained from ABDA: 17
Interactions obtained from DrugBank: 16
Correlating interactions: 7

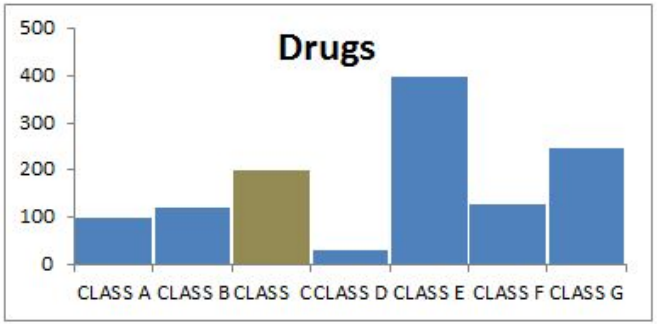
Interaction partner	Language	Datasource
Buprenorphine	EN	ABDA
Carbamazepine	EN	ABDA
Cimetidine	EN	Correlative
Clozapine	EN	Correlative
Disulfiram	EN	ABDA
Esomeprazole	EN	ABDA
Ethanol	EN	ABDA
Fluconazole	EN	Correlative
Indinavir	EN	DrugBank
Itraconazole	EN	Correlative
Kava	EN	DrugBank
Ketoconazole	EN	Correlative
Mephenytoin	EN	ABDA
Miconazole	EN	ABDA
Nelfinavir	EN	DrugBank
Omeprazole	EN	Correlative
Phenytoin	EN	ABDA
Posaconazole	EN	ABDA
Ritonavir	EN	DrugBank
Saquinavir	EN	DrugBank
St. John's Wort	EN	DrugBank
Telithromycin	EN	DrugBank
Tipranavir	EN	DrugBank
Triproldine	EN	DrugBank
Valproic Acid	EN	ABDA
Voriconazole	EN	Correlative

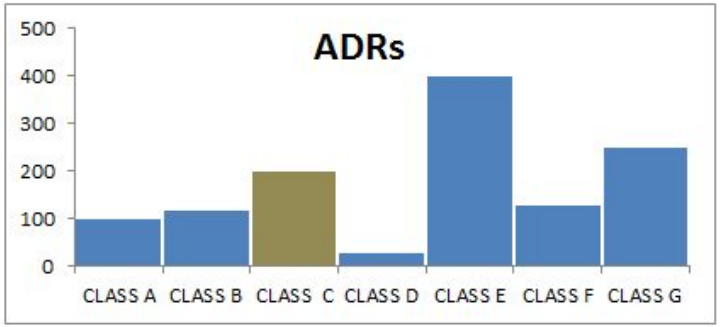
Meeting

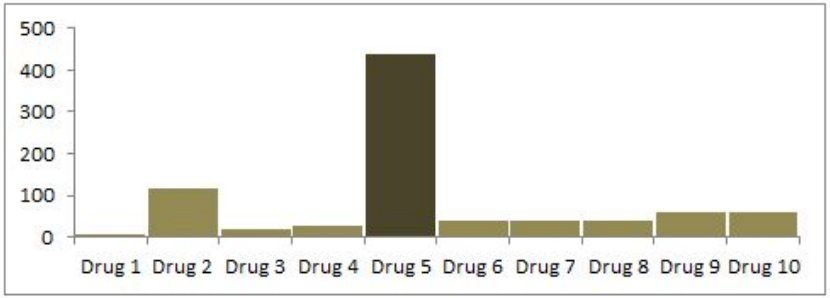
Nov 10, 2016

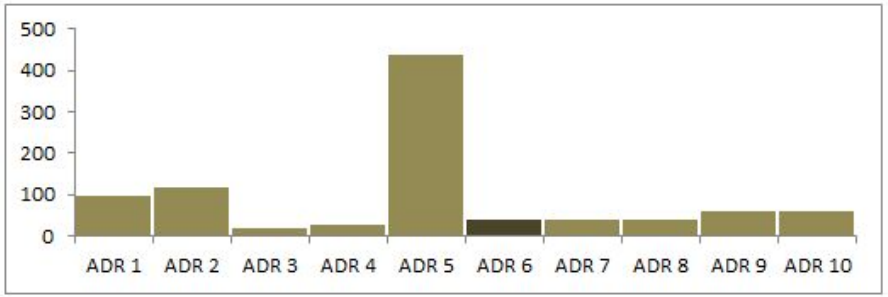
Notes

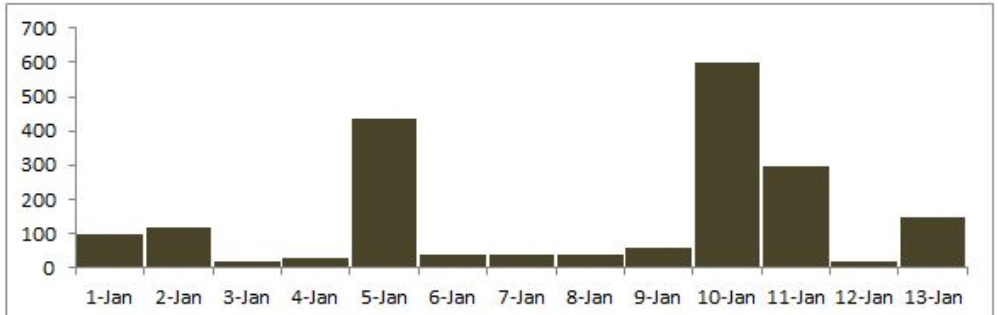
- Can use RanbowBoxes for drug comparisons as each safety evaluation team is assigned a class of drugs, so they most probably have similar ADRs
- Can use Cross filter to show the trends of a specific drug (Next Slide)
  - http://square.github.io/crossfilter/
-


Drugs
500
400
300
200
100
0
CLASS A
CLASS B
CLASS C
CLASS D
CLASS E
CLASS F
CLASS G


ADRs
500
400
300
200
100
0
CLASS A
CLASS B
CLASS C
CLASS D
CLASS E
CLASS F
CLASS G


500
400
300
200
100
0
Drug 1
Drug 2
Drug 3
Drug 4
Drug 5
Drug 6
Drug 7
Drug 8
Drug 9
Drug 10


500
400
300
200
100
0
ADR 1
ADR 2
ADR 3
ADR 4
ADR 5
ADR 6
ADR 7
ADR 8
ADR 9
ADR 10


700
600
500
400
300
200
100
0
1-Jan
2-Jan
3-Jan
4-Jan
5-Jan
6-Jan
7-Jan
8-Jan
9-Jan
10-Jan
11-Jan
12-Jan
13-Jan